NUMERICAL APPROXIMATION OF TIME DEPENDENT FRACTIONAL DIFFUSION WITH DRIFT: NUMERICAL ANALYSIS AND APPLICATIONS TO SURFACE QUASI-GEOSTROPHIC DYNAMICS AND ELECTROCONVECTION

A Dissertation

by

PENG WEI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Andrea Bonito
Committee Members,	Jean-Luc Guermond
	Anastasia Muliana
	Jay Walton
Head of Department,	Emil Straube

August 2019

Major Subject: Mathematics

ABSTRACT

In this work, we approximate a time-dependent problem with drift involving fractional powers of elliptic operators. The numerical scheme is based on an integral representation of the stationary problem at each time step. The integral representation is further approximated by an exponentially-convergent sinc quadrature. This results in multiple independent reaction-diffusion problems approximated using the finite element method. The resulting error between the solution and its approximation in the energy norm is based on a Strang's lemma for the consistency errors generated by sinc quadrature and finite element approximations. The L^2 error is obtained by a standard duality argument. A forward Euler method is considered for the time stepping. It is stable provided the sinc quadrature stepping and time stepping is taken sufficiently small. Under the same condition, we also deduce its first order convergence in time.

We challenge the analyzed numerical scheme in the context of surface quasi-geostrophic (SQG) dynamics and electroconvection system. In each setting, the governing equations are derived from conservations of various physical quantities.

In SQG dynamics the drifting velocity involves the solution to another fractional elliptic problem. The simulations consider two scenarios: inviscid (no diffusion) and inviscid-limit (small diffusion). In both scenarios our simulation results are compared with existing results and good agreements are observed.

In electroconvection, the liquid is located in between two concentric circular electrodes which are either assumed to be of infinite height or slim. Each configuration results in a different nonlocal electro-magnetic model defined on a two dimensional bounded domain. Our numerical simulations indicate that slim electrodes are favorable for electroconvection to occur and are able to sustain the phenomena over long period of time. Furthermore, we provide a numerical study on the influence of the three main parameters of the system: the Rayleigh number, the Prandtl number and the electrodes aspect ratio.

DEDICATION

I dedicate this dissertation to my parents and parents-in-law. A special thanks to my wife Min Jia for all her love and support.

ACKNOWLEDGMENTS

I own my deepest gratitude to my advisor Dr. Andrea Bonito for his guidance, patience, and support through the years. I have been very much impressed and influenced by his extraordinary insights, persistence and high standards in scientific research. I am also grateful to Dr. Jay Walton, Dr. Jean-Luc Guermond, and Dr. Anastasia Muliana, who served as members in my graduate research committee for their valuable insights and contributions to this work.

Special thanks to Dr. Wenyu Lei for intensive discussions on both analysis and simulation aspects of this dissertation. Thanks to Justin Owen for permanent interests in my research and English writing guidance. Appreciation is also expressed to Dr. Wolfgang Bangerth and Dr. Spencer Patty for their support of my numerical simulations using the deal.ii finite element library.

Dr. Alan Demlow taught me theory of finite element methods and adaptive finite element methods. Dr Jean-Luc Guermond taught me advanced topics on finite elements. Dr. Edriss S. Titi taught me numerical analysis in Navier-Stokes equations. I would to thank them and all other professors and teaching assistants who taught me at Texas A&M University. I wish to extend my thanks to the numerical analysis research group at mathematics department, my friends, colleagues, the department staffs for making my experience at Texas A&M University the most treasured memory in my life.

Finally, I am especially grateful to my wife Min Jia for her constant companion and her efforts of keeping my life balanced. I could have never accomplished and experienced this journey without her. I thank my parents and parents in law for their constant love, encouragement and support.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Andrea Bonito (chair), Jay Walton, and Jean-Luc Guermond of the Department of Mathematics and Professor Anastasia Muliana of the Department of Mechanical Engineering.

Part of this work, especially Chapters 3, 4, and 6 is adapted from joint work with Professors Joseph E. Pasciak and Andrea Bonito. All other work conducted for the thesis (or) dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Department of Mathematics, Texas A&M University, the TAMU Hagler Institute for Advanced Study HEEP Graduate Fellowship, and a research fellowship from National Science Foundation through grants DMS-1254618 and DMS-1817691.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

The fractional diffusion has recently attracted a lot of attention. To name a few applications, fractional diffusion has been applied to scientific areas such as finance [7, 8], flow dynamics [9, 10], peridynamics, image processing [11, 12], and others.

1.1 Motivation

The direct motivation of this dissertation stems from the surface quasi-geostrophic dynamics [13] and electroconvection equation [6], the two systems we will study in-depth in this work. In both of these two models, one can realize that the model reductions involving the Dirichlet-to-Neumann (DTN) mapping will introduce fractional Laplacian with power $1/2$. Consequetly, the numerical study of partial differential equations involving fractional differential operators is necessary.

Throughout this work, two different types of fractional differential operators are considered: the *spectral* fractional elliptic operator and the *integral* fractional Laplacian. The former is defined by the spectral decomposition. Suppose X and Y are both Hilbert spaces with Y being dense and compactly embedded in X. Suppose $a(\cdot,\cdot)$ is a bilinear form defined on $Y \times Y$ which is coercive, bounded and symmetric, the unbounded linear operator $L : Y \to X$ can be derived from the bilinear form, i.e.,

$$\langle Lu, v\rangle = a(u, v), \qquad \text{for all } v \in Y.$$

Here $\langle\cdot,\cdot\rangle$ denotes the duality pairing. The spectral fractional operator L^s for $s \in (0, 1)$ is defined by an eigenvalue shift of the eigenfunction expansion, namely, given $v \in D(L^s) := \{v \in L^2(\Omega) : L^s v \in L^2(\Omega)\}$,

$$L^s v = \sum_{j=1}^{\infty} \lambda_j^s (v, \psi_j)\psi_j, \tag{1.1}$$

where $(\cdot,\cdot)$ denotes the X inner product and $\{\psi_j\}$ is an X-orthonormal basis of eigenfunctions of L corresponding to positive eigenvalues $\{\lambda_j\}$ which are organized in the ascending order. For a

Lipschitz domain Ω, we consider $Y = H_0^1(\Omega)$ and $X = L^2(\Omega)$ When $L = -\Delta$, L^s is referred to as the *spectral fractional Laplacian*. In addition, when Ω is convex, we have $D(L) = H_0^1(\Omega) \cap H^2(\Omega)$.

The second type of fractional operator, the integral fractional Laplacian, is defined through the Fourier transform on the whole space $\mathbb{R}^d$. For v in the Schwartz space, the integral fractional Laplacian is a pseudo-differential operator with symbol $|\zeta|^{2s}$,

$$\mathscr{F}((-\Delta)^s v)(\zeta) = |\zeta|^{2s} \mathscr{F}(v)(\zeta), \tag{1.2}$$

with $\mathscr{F}$ denoting the Fourier transform

$$\mathscr{F}(v)(\zeta) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-i\zeta\cdot x} v(x)\ dx.$$

The definition can be extended to the Sobolev space $\mathbb{H}^s(\mathbb{R}^d)$ by a density argument.

We comment that the two types of fractional operators are different. In fact, it is shown in [14, Theorems 1-2] that their difference is positivity preserving and positive definite.

The model problem we consider in this dissertation is the following time dependent equation with drift and involving spectral fractional operators:

$$\frac{\partial}{\partial t} q + \boldsymbol{u} \cdot \nabla q + \kappa L^s q = f \qquad \text{in} \quad \Omega \times (0, \mathtt{T}), \tag{1.3}$$

$$q = v \qquad \text{on} \quad \Omega \times \{t = 0\}. \tag{1.4}$$

Here the domain $\Omega \subset \mathbb{R}^d$ ($d = 2, 3$) is bounded and is with Lipschitz boundary, the coefficient $\kappa \geq 0$ is taken as constant, the final time $\mathtt{T} > 0$, the fractional power $s \in [1/2, 1)$, the right hand side $f \in L^2(0, \mathtt{T}; L^2(\Omega))$, and the initial condition $v \in L^2(\Omega)$. We assume that the velocity $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ belongs to $\boldsymbol{U}$, where

$$\boldsymbol{U} := \left\{ \boldsymbol{v} \in \mathbf{L}^\infty(\Omega) := L^\infty(\Omega)^d \ :\ \nabla\cdot\boldsymbol{v} = 0 \text{ in } \Omega,\ \boldsymbol{v}\cdot\boldsymbol{n} = 0 \text{ on } \partial\Omega \right\}, \tag{1.5}$$

with $\boldsymbol{n}$ the outward pointing vector normal to $\partial\Omega$. We remark that in surface quasi-geostrophic and electroconvection systems, the velocity $\boldsymbol{u}$ and q are fully coupled.

Equation (1.3) is supplemented with vanishing boundary condition $q = 0$ on $\partial\Omega \times [0, \mathtt{T}]$. We also discuss the periodic boundary condition for $\Omega = (0,1)^d$ and $L^s = (-\Delta)^s$ the spectral fractional Laplacian. In this case, the solution q is further fixed by a mean value condition $\int_\Omega q = 0$ for $t \in [0, \mathtt{T}]$.

1.2 Existing numerical results for fractional operators

In this section we review some existing approaches to the approximation of the following stationary model problem. For given $f \in L^2(\Omega)$ and $s \in (0,1)$, find $u \in D(L^s)$ such that

$$L^s u = f \qquad \text{in} \quad \Omega, \tag{1.6}$$

with vanishing Dirichlet boundary conditions (understood in the sense that $u = 0$ on $\partial\Omega$ for spectral case, and $u = 0$ on $\mathbb{R}^d \backslash \Omega$ for integral case). Here L^s is either spectral fractional operator defined by (1.1) or the integral fractional laplacian defined by (1.2).

Because fractional operators are nonlocal, applying standard finite element method directly to (1.6) results in a dense computationally expensive system matrix. Several alternatives have been proposed in the literature.

One natural approach to approximate the spectral fractional operator is to utilize the definition (1.1), where L is replaced by a finite dimensional approximation L_h,cf. [15, 16, 1, 2]. The disadvantage of this approach is the demanding computational cost. Indeed, in order to obtain an acceptable approximation, it requires the computation of all the eigenfunctions and eigenvalues of a large-scaled system matrix.

Caffarelli and Silvestre [20] showed the equivalence between a fractional Laplacian problem defined on $\mathbb{R}^d$ and an extended problem on the upper half space $\mathbb{R}^d \times (0, \infty)$. The extended problem

is a mix-boundary value problem involving only full operators:

$$\nabla\cdot(y^\alpha\nabla\mathscr{U}) = 0 \quad \text{in } \mathbb{R}^d \times (0,\infty), \quad \mathscr{U}(x,0) = u(x) \quad \text{on } \mathbb{R}^d \times \{0\}, \tag{1.7}$$

where $\alpha = 1 - 2s \in (-1, 1)$. The relation between (1.6) and (1.7) is

$$c_s(-\Delta)^s u(x) = -\lim_{y\to 0^+} y^\alpha \mathscr{U}_y(x,y), \tag{1.8}$$

where the fractional operator is of integral type (1.2), and $c_s := 2^{1-2s}\Gamma(1-s)/\Gamma(s)$ is a normalization constant with $\Gamma(\cdot)$ the gamma function, i.e., for $\Re(z) > 0$,

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\,dx.$$

Their result was further extended by Stinga and Torrea in [21] to the problem imposed on the bounded domain $\Omega \subset \mathbb{R}^d$. The extended problem is now posed on the semi-infinite cylinder $\mathcal{C} := \Omega \times (0,\infty) \subset \mathbb{R}^{d+1}$,

$$\nabla\cdot(y^\alpha\nabla\mathscr{U}) = 0 \quad \text{in } \mathcal{C}, \quad \mathscr{U} = 0 \quad \text{on } \partial_L\mathcal{C}, \quad \mathscr{U}(x,0) = u(x) \quad \text{on } \Omega \times \{0\}, \tag{1.9}$$

where $\partial_L\mathcal{C} := \partial\Omega \times (0,\infty)$ is the lateral boundary of $\mathcal{C}$. For equations (1.6) and (1.9), the same relation (1.8) holds but with $(-\Delta)^s$ the spectral fractional type (1.1).

By extending to the $\mathbb{R}^{d+1}$ space, the partial differential equations (1.7) and (1.9) are well-posed in weighted Sobolev spaces. The finite element approximation is restricted to a truncated domain $\mathcal{C}_\mathcal{Y} := \Omega \times (0,\mathcal{Y})$ for certain $\mathcal{Y} \in (1,\infty)$, thanks to the exponential decay property of $\mathscr{U}$. Notice that the mesh along the extended direction needs to be graded geometrically towards the $\mathbb{R}^d$ plane in order to capture the singular behavior of the solution $\mathscr{U}$ on the extended variable y.

The last approach originates in the Dunford-Taylor representation, see Kato [22]. This is the starting point of the numerical methods proposed by Bonito et. al. [**?** 25, 26]. We begin with the

spectral fractional operator, which admits the following integral representation

$$L^{-s}f = \frac{1}{2\pi i}\int_{\mathcal{C}} z^{-s}(zI-L)^{-1}f\,dz \quad \text{for } f \in L^2(\Omega). \tag{1.10}$$

Here $\mathcal{C}$ is a complex Jordan curve running in the resolvent set of L, and $z^{-s} = e^{-s\ln z}$ with the branch cut for the complex logarithm along the negative direction of real axis. We remark that (1.10) is the definition of negative powers of regularly accretive operator. A proper deformation of contour $\mathcal{C}$ yields the following so-called Balakrishnan formula, c.f. [23],

$$u = L^{-s}f = \frac{\sin(\pi s)}{\pi}\int_0^\infty \mu^{-s}(\mu I + L)^{-1}f\,d\mu. \tag{1.11}$$

Following the above infinite integral formula (1.11), the numerical approximation consists of two steps: (i) apply a numerical integral scheme to the right-hand side of (1.11) on a set of quadrature points $\{\mu_j\}$ and truncate the higher order terms; (ii) use finite element method to approximate $(\mu_j I + L)^{-1}f$ for each j on the *same* mesh. The first step is achieved by applying an exponentially convergent sinc quadrature scheme [24], see [23, 25, 26]. The truncation of higher order terms is a consequence of the exponentially decay property of the integrand. For simplicity, a conforming piecewise linear finite element space is applied in the second step.

The direct representation (1.10) or (1.11) cannot be carried onto the integral fractional laplacian (1.2) because neither (1.10) or (1.11) are well-defined for integral fractional laplacian. However, a Balakrishnan approach is possible upon using Fourier transform and Parserval's identity. This results in a weak formulation to (1.6): find $u \in \tilde{H}^s(\Omega)$, such that

$$((-\Delta)^s\tilde{u}, (-\Delta)^s\tilde{\varphi})_{\mathbb{R}^d} = \frac{\sin(\pi s)}{\pi}\int_{-\infty}^{\infty} e^{-2st}\left(e^{2t}(-\Delta)(I-e^{2t}\Delta)^{-1}\tilde{u},\varphi\right)_\Omega\,dt = \int_\Omega f\varphi\,dx,$$

for all $\varphi \in \tilde{H}^s(\Omega)$. Here $\tilde{u}$ denotes extension by zero, and $\tilde{H}^s(\Omega)$ denotes the space of functions in Ω whose extension by zero belongs to the Sobolev space $H^s(\mathbb{R}^d)$. A conjugate-gradient type method was developed to solve the weak problem, see [27]. This in turn requires the approximated

evaluation of the bilinear form $((-\Delta)^s\tilde{u}, (-\Delta)^s\tilde{\varphi})_{\mathbb{R}^d}$. In contrast and comparison to the approximation scheme for spectral fractional operators, the numerical approximation to the evaluation of $((-\Delta)^s\tilde{u}, (-\Delta)^s\tilde{\varphi})_{\mathbb{R}^d}$ takes three steps: (i) apply a numerical integral scheme on finitely many quadrature points $\{t_j\}$ so that the infinite integral can be approximated by a finite sum of inner products involving the solution to an elliptic partial equation $(-\Delta)(I - e^{2t}\Delta)^{-1}\tilde{u}$ on the entire space $\mathbb{R}^d$; (ii) on each of the quadrature points t_j, approximate the problem on the entire space by a problem on a truncated bounded convex domain with vanishing boundary condition; (iii) use continuous piece-wise linear finite element scheme to find the approximated solution for each of the truncated problems.

We refer to the review paper [28] on a thorough discussion of the two methods and their extensions, including time dependent problems, obstacle problems and adaptive finite elements.

1.3 Contents of this dissertation

This work consists of numerical analysis of the approximation scheme to the problem (1.3), together with its applications to the surface quasi-geostrophic dynamic and the electroconvection system. A general description of the contents of this dissertation is given as follows.

Approximation to the stationary problem

The numerical approximation to the stationary problem

$$\boldsymbol{u} \cdot \nabla w + L^s w = f \quad \text{in } \Omega, \qquad w = 0 \quad \text{on } \partial\Omega, \tag{1.12}$$

for $s \in [1/2, 1)$ and $f \in L^2(\Omega)$ is shown to be fundamental to the numerical analysis of the time-dependent problem (1.3).

For the simple case $\boldsymbol{u} = 0$ the regularity of the solution w to equation (1.12) follows directly from the definition of L^s. The regularity of the solution for general case is discussed in Proposition 3.1. The construction of numerical algorithm begins with the derivation of a Balakrishnan-type integral representation for the weak formulation of the stationary problem (1.12). Thanks to the integral representation, a sinc quadrature scheme followed by a first-order finite element approxi-

mation is proposed in Section 3. The consistency error between the original bilinear form and its finite element approximation is illustrated in Theorem 3.8. It consists of the sinc quadrature error and the finite element approximation error. A second Strang's lemma is advertised to derive the $\mathbb{H}^s(\Omega)$-error between the solution w and its finite element approximation w_h, which is the topic of Lemma 3.11. We finally apply the standard duality argument with the assistant of the regularity result Proposition 3.1 to obtain the $L^2(\Omega)$-error estimate, see Theorem 3.13.

Approximation to the time dependent problem

The numerical scheme to the time dependent problem (1.3) distinguishes two cases: the approximation to the homogeneous problem for $f = 0$ and $v \neq 0$, and the approximation to the non-homogeneous problem for $f \neq 0$ and $v = 0$. In both cases the numerical scheme consists of two steps: the finite element approximation to static problem and discretization in time domain.

We begin with the homogeneous problem. We denote $\mathcal{A}$ to be the inverse of the solution operator to the weak stationary problem, and $\mathcal{A}_h$ the corresponding finite element approximated operator such that the approximating solution admits a representation $q_h := e^{-t\mathcal{A}_h}\pi_h v$ with π_h the L^2 projection onto V_h. Inspired by the techniques developed in [29], both $q := e^{-t\mathcal{A}}v$ and q_h can be represented with the Dunford-Taylor integral, e.g.,

$$q(x;t) = \frac{1}{2\pi i}\int_{\mathcal{C}} e^{-tz}(zI - \mathcal{A})^{-1}v dz, \tag{1.13}$$

where the curve $\mathcal{C}$ runs in the resolvent set of $\mathcal{A}$. The integral representation of q_h is similar, with $\mathcal{A}$ and v replaced by $\mathcal{A}_h$ and $\pi_h v$ respectively. With the two integral representations we show in Lemma 4.3 that for a fixed time $t > 0$, the $L^2(\Omega)$ error estimate for the finite element approximation $q_h(t)$ is the sum of the exponentially converging sinc approximation error and the optimal-rate finite element error, with a multiplicative constant depending on t, α, and the regularity of initial data v.

A forward Euler time stepping is advocated for the time discretization. We first show that the time stepping is stable under the condition $\tau \leq Ch^{2s}$, where τ is the time step; see Lemma 4.1. Following the techniques developed in [30], Lemma 4.6 guarantees a first-order convergence rate

in time. The fully discretization error is obtained by combining the error estimates from space and time discretizations. The result is presented in Theorem 4.7.

We also consider the non-homogeneous problem. Thanks to Duhamel's principle, for the finite element approximation, we are able to obtain results similar to the homogeneous case; see Lemma 4.8. It is worth mentioning that the rate of convergence for the space approximation will degenerate for not sufficiently smooth data f.

We employ again the forward Euler time stepping fo the time discretization. The stability requirement is the same as in the homogeneous case. Lemma 4.9 provides the first order convergence rate in time. The error between the solution to the original problem (1.3) and its fully discretized approximation consists of the errors from sinc approximation, together with the space and time discretization; see Theorem 4.10.

Simulation of the surface quasi-geostrophic dynamic

The quasi-geostrophic theory has been a successful model for the study of oceanic and atmospheric dynamics in the mid-to-high latitude region of the earth where the Coriolis effect is significant. The characteristic property of the quasi-geostrophic system is the conservation of potential vorticity along the geostrophic flow; see [31, 32, 33]. If in addition to the potential vorticity conservation, the assumption on the surface buoyancy (or potential temperature) conservation reduces the model to the surface quasi-geostrophic equation (SQG), cf. [13, 34]. In this work we will consider the following generalized SQG equation:

$$\frac{\partial}{\partial t}\theta + \boldsymbol{u}\cdot\nabla\theta + \kappa(-\Delta)^s\theta = 0, \quad \boldsymbol{u} = \nabla^\perp\Psi := (-\frac{\partial}{\partial y}\Psi, \frac{\partial}{\partial x}\Psi), \quad (-\Delta)^{1/2}\Psi = -\theta, \quad (1.14)$$

all in $\Omega\times(0,\mathtt{T})$, with the fractional power $s\in[1/2,1)$, the constant $\kappa\geq 0$, and the initial condition $\theta(\boldsymbol{x},0)=\theta_0(\boldsymbol{x})\in L^2(\Omega)$. Here both the two fractional operators in (1.14) are spectral fractional Laplacians as defined by (1.1). The temperature potential $\theta:\Omega\times[0,\mathtt{T}]\to\mathbb{R}$ is coupled with periodic boundary condition and is additionally fixed by the averaging constraint $\int_\Omega\theta\,dx=0$. The vector field $\boldsymbol{u}$ represents the velocity of the fluid, and $\Psi:\Omega\times[0,\mathtt{T}]\to\mathbb{R}$ is a stream function

representing the geopotential.

As indicated by the definition of the velocity $\boldsymbol{u} := \nabla^{\perp}\Psi$, the motion direction of the fluid is always along the iso-bars. This is due to the fact that the gradient pressure is balanced by the Coriolis force of the earth. For instance, for a high-pressure atmosphere system in the southern hemisphere, the pressure gradient is pointing outward, while the Coriolis force is perpendicular to the direction of the velocity to the left side. As a consequence, viewing from above, the rotation of atmosphere should be counter-clockwise in order to balance the two forces. The fractional dissipation term $\kappa(-\Delta)^s\theta$ represents the so-called Ekman pumping, which depicts a balance between the vertical component of the Coriolis force and the vertical frictional force. The detailed derivation of the governing equations (1.14) are proposed in Section 5.1.

The numerical approximation scheme follows the essence of the approaches described in Chapter 4 with two additional techniques. For the time discretization, we apply the second-order two-stage explicit strongly stability preserving Runge-Kutta method (SSP-RK2) proposed in [35]. It allows a second order convergence rate in time. In addition, due to the well-known fact that when applying continuous finite element method to the transport equations, oscillations may arise. The defect is circumvented by applying a second-order maximum principle preserving artificial viscosity, as proposed in [36].

We run simulations on two settings depending on the choice of κ: the inviscid case $\kappa = 0$, and the viscid-limit case $0 < \kappa \ll 1$. Two smooth initial functions are considered, and the corresponding solutions are compared with existing results in [1] and [2].

Simulation of the electroconvection equations

The electroconvection model describes the convective flow in thin layered liquid crystals. The liquid is located in between two concentric circular electrodes, which are assumed either to be of infinite heigh (infinite case) or zero height (slim case). The two configurations result in spectral fractional laplacian or integral fractional laplacian respectively in the electromagnetic models defined on a two-dimensional annular domain.

The governing equations contain the incompressible Navier-Stokes equation for the fluid ve-

locity and pressure, the equation for the charge density, and the equation for the electric potential in the two dimensional liquid region. The resulting numerical method consists of approximating the above-mentioned quantities. Finite element methods for the space discretization coupled with standard time stepping methods are put forward. For the same reasons as in the simulation of SQG equations, in the numerical algorithm we additionally incorporate the SSP-RK2 scheme for time discretization, and the second-order maximum principle preserving artificial viscosity for stabilization.

Our key objectives of the simulations are the assessment of the advantages of the two electrode configurations and the effect of three critical non-dimensional parameters emerged in the electroconvection process. The three control parameters are the Rayleigh number $\mathcal{R}$ representing the effect between the electric forcing and viscous dissipation on the charged fluid, the Prandtl number $\mathcal{P}$ measuring the relaxation ability between the charge relaxation and the fluid viscous relaxation, and the boundary aspect ratio α characterizing the geometry of the fluid domain.

Our simulation results are compared with existing results in [6, 37]. In particular, our numerical simulations reveal that the slim electrodes configuration is more favorable than the infinite model for electroconvection: it requires less energy and possesses the capability of long-term sustainable convection. Moreover, we find that the Prandtl number does not effect the long-term behavior of the charge density distribution and the fluid dynamics. In contrast, for various aspect ratios, we determine the critical Rayleigh numbers above which the electroconvection occurs. The number of pairs of counter-rotating convective flows are also determined from our observations on the simulation results of the velocity fields.

2. PRELIMINARIES

In this chapter we review the fundamental tools that will be utilized throughout this work. It is outlined as follows. We introduce commonly used notations and the L^p spaces in Section 2.1. The Sobolev spaces and their equivalent real interpolation definitions are overviewed in Section 2.2. We put forward the definition of dotted spaces in Section 2.3 and the definition of fractional powers of elliptic operators in Section 2.4. The two major numerical approximation methodologies, the Galerkin finite elements and sinc quadrature scheme are advocated in Sections 2.5 and 2.6 respectively. We finish this chapter by reviewing the numerical algorithms for the approximations of spectral fractional Laplacians in Secrtion 2.7 and integral fractional Laplacians in Section 2.8.

2.1 Notations

We denote by c and C a generic positive constant which may change in various occasions. We sometimes use $A \precsim B$ to represent $A \leq cB$ when c is independent of A and B. Throughout this work the domain, often denoted by Ω, is a nonempty open set in the n-dimensional real Euclidean space $\mathbb{R}^d$. We denote by $\partial\Omega$ its boundary in $\mathbb{R}^{d-1}$.

The norm of a bounded operator $F : X \to Y$ between two Banach spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ is denoted by

$$\|F\|_{X\to Y} := \sup_{u\in X, \|u\|_X=1} \|Fu\|_Y ,$$

and we write $\|F\| := \|F\|_{L^2(\Omega)\to L^2(\Omega)}$ for the particular case when $X = Y = L^2(\Omega)$.

Given $1 \leq p \leq \infty$ and a Lebesgue measurable function $u : \Omega \to \mathbb{R}$, we define the L^p-norm $\|u\|_{L^p(\Omega)}$ by

$$\|u\|_{L^p(\Omega)} = \left(\int_\Omega |u(x)|^p \ dx\right)^{1/p}, \text{ for } p \in [1, \infty), \text{ and}$$
$$\|u\|_{L^\infty(\Omega)} = \operatorname*{ess\,sup}_{x\in\Omega} |u(x)| , \text{ for } p = \infty.$$

We denote by $L^p(\Omega)$ the class of all measurable functions u for which the norm $\|\cdot\|_{L^p(\Omega)}$ is finite.

Without ambiguity, we denote $\|u\| := \|u\|_{L^2(\Omega)}$. Also we denote the L^2 inner product by

$$(u, v) := (u, v)_{L^2(\Omega)} = \int_\Omega u\bar{v}\, dx,$$

where $\bar{v}$ stands for the complex conjugate of v. For a vector-valued function $\boldsymbol{v} : \Omega \to \mathbb{R}^d$, we use the notation $\mathbf{L}^\infty(\Omega) := L^\infty(\Omega)^d$ with corresponding norm $\|\boldsymbol{v}\|_{\mathbf{L}^\infty(\Omega)} := \||\boldsymbol{v}|\|_{L^\infty(\Omega)}$.

For $\mathtt{T} > 0$ and $1 \le p \le \infty$, we consider $u : \Omega \times [0, \mathtt{T}] \to \mathbb{R}$ as a function of t with values in a Banach space X. The space $L^p(0, \mathtt{T}; X)$ consists of all X-valued functions whose norm is in $L^p(0, \mathtt{T})$. It is a Banach space when equipped with norm

$$\|u\|_{L^p(0,\mathtt{T};X)} = \left(\int_0^{\mathtt{T}} \|u(t)\|_X^p \ \ dt\right)^{1/p}, \text{ for } p \in [1, \infty), \text{ and}$$

$$\|u\|_{L^\infty(0,\mathtt{T};X)} = \operatorname*{ess\,sup}_{t\in(0,\mathtt{T})} \|u(t)\|_X, \text{ for } p = \infty.$$

Let $\alpha = (\alpha_1, \cdots, \alpha_d)$ be a d-tuple consists of non-negative integers, we call α a *multi-index* and denote its degree $|\alpha| := \sum_{j=1}^d \alpha_j$. We denote by $x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, and similarly denote the differential operator

$$D^\alpha := D_1^{\alpha_1} \cdots D_d^{\alpha_d},$$

where $D_j = \partial/\partial x_j$, with the convention $D^{(0,\cdots,0)}u = u$. For any nonnegative integer m let C_0^m be the space consists of all functions ϕ for which ϕ itself together with all its partial derivatives $D^\alpha\phi$ of orders $|\alpha| \le m$ are continuous and have compact support on Ω. Let $C_0^\infty(\Omega) := \cap_{m=0}^\infty C_0^m(\Omega)$.

A function $f : \Omega \to \mathbb{R}$ is said to be locally integrable if f is integrable on each compact subset $\Omega_c \subset \Omega$. The set of all locally integrable functions on Ω is denoted $L^1_{loc}(\Omega)$.

2.2 Sobolev spaces

In this work, we employ Sobolev spaces to characterize the smoothness of functions, in particular, the smoothness of given data and the solutions to the equations. Some essential concepts and properties of Sobolev spaces are summarized in this section. We refer to [38, 39] for more detailed discussions on Sobolev spaces.

The weak derivative

A function $u \in L^1_{loc}(\Omega)$ has a *weak derivative* $D^\alpha u =: v_\alpha$ provided $v_\alpha \in L^1_{loc}(\Omega)$ and satisfies

$$\int_\Omega u(x) D^\alpha \phi \, dx = (-1)^{|\alpha|} \int_\Omega v_\alpha \phi \, dx, \qquad \text{for all } \phi \in C_0^\infty(\Omega).$$

It can be easily checked through direct verification of the above identity that for u being sufficiently smooth to ensure a continuous partial derivative $D^\alpha u$ in the classical sense, then $D^\alpha u$ is also a weak derivative. Therefore, unless explicitly declaired, throughout this work the partial derivative D^α denotes the weak derivative.

Sobolev spaces

Given an integer $m \geq 0$, and given $p \geq 1$, the integer-ordered Sobolev space $W^{m,p}(\Omega)$ is defined by

$$W^{m,p}(\Omega) := \{u \in L^p(\Omega) \ : \ D^\alpha u \in L^p(\Omega) \ \text{ for all } \ |\alpha| \leq m\}.$$

It is equipped with semi norm $|\cdot|_{W^{m,p}(\Omega)}$ and norm $\|\cdot\|_{W^{m,p}(\Omega)}$

$$|u|_{W^{m,p}(\Omega)} = \Big(\sum_{|\alpha|=m} \|D^\alpha u\|_{L^p(\Omega)} \Big)^{1/p}, \quad \text{ and } \ \|u\|_{W^{m,p}(\Omega)} = \Big(\sum_{|\alpha|\leq m} \|D^\alpha u\|_{L^p(\Omega)} \Big)^{1/p}.$$

Given $r \in (0,1)$, the fractional-ordered Sobolev space $W^{r,p}(\Omega)$ is defined by

$$W^{r,p}(\Omega) := \big\{u \in L^p(\Omega) \ : \ |u|_{W^{r,p}(\Omega)} < \infty\big\},$$

where the seminorm is the so-called Aronszanjn-Slobodeckij seminorm

$$|u|_{W^{r,p}(\Omega)} = \Big(\iint_{\Omega\times\Omega} \frac{|u(x)-u(y)|^p}{|x-y|^{d+rp}} \, dx \, dy \Big)^{1/p},$$

here d corresponds to the dimension of the space $\mathbb{R}^d$. Sobolev spaces of fractional order greater than one are defined as follows. For $r > 1$ not an integer, the unique decomposition $r = m + \sigma$

with $\sigma \in (0,1)$ and $m \in \mathbb{N}_+$ allows us to define $W^{r,p}(\Omega)$ by

$$W^{r,p}(\Omega) := \left\{ u \in W^{m,p}(\Omega) \ : \ |D^\alpha u|_{W^{r,p}(\Omega)} < \infty \ \text{ for all } \alpha \ \text{ s.t. } \ |\alpha| = m \right\}.$$

Its corresponding full norm is given by

$$\|u\|_{W^{r,p}(\Omega)} = \Big(\sum_{|\alpha|=m} |D^\alpha u|^p_{W^{r,p}(\Omega)} + \|u\|^p_{W^{m,p}(\Omega)} \Big)^{1/p}.$$

For $r \geq 0$, $W^{r,p}(\Omega)$ equipped with full norm $\|\cdot\|_{W^{r,p}(\Omega)}$ is a Banach space (cf. [38]). In particular, when $p = 2$, $H^r(\Omega) := W^{r,2}(\Omega)$ is a Hilbert space.

We denote by $H^1_0(\Omega)$ the closure of $C^\infty_0(\Omega)$ in $H^1(\Omega)$. We immediately remark that $H^1_0(\Omega)$ can be characterized by

$$H^1_0(\Omega) := \left\{ u \in H^1(\Omega) \ : \ u = 0 \text{ on } \partial\Omega \right\},$$

and after invoking Poincare's inequality, the seminorm $|\cdot|_{H^1(\Omega)}$ is equivalent to the full norm $\|\cdot\|_{H^1(\Omega)}$. In the remaining of this work, unless explicitly declared, the $H^1_0(\Omega)$ will always be equipped with the (semi)norm $|\cdot|_{H^1(\Omega)}$. We define the dual space of $H^1_0(\Omega)$, denoted by $H^{-1}(\Omega)$ to be the collection of all bounded linear functionals F acting on $H^1_0(\Omega)$ such that the operator norm

$$\|F\|_{H^{-1}(\Omega)} := \sup_{0 \neq v \in H^1_0(\Omega)} \frac{\langle F, v \rangle}{\|v\|_{H^1_0(\Omega)}} < \infty.$$

Here $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $H^{-1}(\Omega)$ and $H^1_0(\Omega)$. It is easy to observe that $L^2(\Omega) \subset H^{-1}(\Omega)$ since in this case $\langle F, v \rangle$ can be trivially identified with (F, v).

Scales of interpolation spaces

Let X and Y be two Banach spaces with Y continuously embedded and dense in X. For any $v \in X$ and real variable $t > 0$, the *K-functional* is defined by

$$K(v,t) := \inf \Big\{ \|x\|_X + t \|y\|_Y \ : \ v = x + y, \ x \in X, \ y \in Y \Big\}. \tag{2.1}$$

with $x \in X$ and $y \in Y$. This allows us to define the intermediate spaces $[X, Y]_{r,p}$

$$[X, Y]_{r,p} := \left\{ v \in X \ : \ \|v\|_{[X,Y]_{r,p}} < \infty \right\},$$

for $0 < r < 1$ and $1 \leq p < \infty$. Here the norm is given by

$$\|v\|_{[X,Y]_{r,p}} = \left(\int_0^\infty t^{-rp} K^p(v, t) \frac{dx}{t} \right)^{1/p}.$$

For convention we denote the corner cases $[X, Y]_{0,p} = X$ and $[X, Y]_{1,p} = Y$. It can be shown in [39] that all the intermediate spaces are Banach, and for $0 \leq r_1 \leq r_2 \leq 1$, we have $[X, Y]_{r_2,p} \subset [X, Y]_{r_1,p}$ with continuous embedding. More importantly, we introduce the following crucial interpolation inequality (cf. [39, Lemma 22.3]).

Proposition 2.1. Let $[X, Y]_{r,p}$ and $[X', Y']_{r,p}$ both be interpolation spaces. Let $\mathcal{L}$ be the linear operator that maps X into X' with $\|\mathcal{L}v\|_{X'} \leq M_1 \|v\|_X$ for all $v \in X$, and maps Y into Y' with $\|\mathcal{L}v\|_{Y'} \leq M_2 \|v\|_Y$ for all $v \in Y$, then for all $v \in [X, Y]_{r,p}$,

$$\|\mathcal{L}v\|_{[X',Y']_{r,p}} \leq M_1^{1-r} M_2^r \|v\|_{[X,Y]_{r,p}}. \tag{2.2}$$

Thanks to the interpolation theory, we now define the interpolation spaces between the Hilbert spaces $H^{-1}(\Omega)$ and $H^2(\Omega) \cap H_0^1(\Omega)$ by

$$\mathbb{H}^r := \mathbb{H}^r(\Omega) := \begin{cases} H_0^1(\Omega) \cap H^r(\Omega), & 1 < r \leq 2, \\ [L^2(\Omega), H_0^1(\Omega)]_{r,2}, & 0 \leq r \leq 1, \\ [H^{-1}(\Omega), L^2(\Omega)]_{1+r,2}, & -1 \leq r < 0, \end{cases} \tag{2.3}$$

We remark that for $r \in [0, 1]$, the k-functional is defined to be

$$K(v, t) := \inf \left\{ \|x\|_{L^2(\Omega)} + t \, |y|_{H^1(\Omega)} \ : \ v = x + y, \ x \in L^2(\Omega), \ y \in H_0^1(\Omega) \right\}. \tag{2.4}$$

The two different definitions of Sobolev spaces are equivalent, namely $\mathbb{H}^r = H^r(\Omega)$ for $r \in [-1, 2]$ with equivalent norms. In this dissertation we shall use both definitions interchangeably.

2.3 Dotted space and its characterization

Let $a_\Omega(\cdot, \cdot)$ be a symmetric, coercive and continuous bilinear form on $H_0^1(\Omega)$, which means there exist two constants c and C, such that

$$a_\Omega(w, w) \geq c \|w\|^2_{H^1(\Omega)}, \quad |a_\Omega(w, v)| \leq C \|w\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}, \quad \forall\, w, v \in H_0^1(\Omega).$$

For a given $f \in L^2(\Omega)$, the Lax-Milgram theory ensures that the problem of finding $w \in H_0^1(\Omega)$ satisfying

$$a_\Omega(w, v) = (f, v) \qquad \text{for all } \; v \in H_0^1(\Omega) \tag{2.5}$$

has a unique solution. This in turn allows us to define the solution operator $T : L^2(\Omega) \to H_0^1(\Omega)$ by $Tf = w$ and further define $L := T^{-1}$ its inverse operator with domain $D(L) := \text{Range}(T)$.

Note that T is compact and symmetric on $L^2(\Omega)$, therefore there exists a complete set of $L^2(\Omega)$-orthonormal eigenfunctions $\{\psi_j\}_{j=1}^\infty$ corresponding to real-valued positive eigenvalues $\infty > \mu_1 \geq \mu_2 \geq \mu_3 \geq \cdots > 0$. We then set $\lambda_j = \mu_j^{-1}$ so that $\{\lambda_j,\, \psi_j\}$ are eigenpairs of the operator L.

For $r > 0$, the *dotted spaces* $\dot{H}^r(\Omega)$ are defined by

$$\dot{H}^r := \dot{H}^r(\Omega) := \left\{ v = \sum_{j=1}^\infty (v, \psi_j)\psi_j \in L^2(\Omega) \;:\; \sum_{j=1}^\infty \lambda_j^r \left|(v, \psi_j)\right|^2 < \infty \right\}.$$

They are Hilbert spaces equipped with the inner product

$$(w, v)_{\dot{H}^r} := \sum_{j=1}^\infty \lambda_j^r (w, \psi_j)(v, \psi_j). \tag{2.6}$$

We denote by $\dot{H}^{-r} := \dot{H}^{-r}(\Omega)$ their dual spaces equipped with the norm

$$\|f\|_{\dot{H}^{-r}} := \Big(\sum_{j=1}^\infty \lambda_j^{-r} \left|\langle f, \psi_j \rangle\right|^2 \Big)^{1/2},$$

where $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_r$ is the duality product between $\dot{H}^r$ and its dual.

We can naturally extend the definition of the operator T to $H^{-1}(\Omega) \to H_0^1(\Omega)$ by setting $TF := w$ when $w \in H_0^1(\Omega)$ is the solution to Equation (2.5) with f replaced by $F \in H^{-1}(\Omega)$ and the inner product replaced by the duality pairing $\langle F, v \rangle$. We require the following instrumental assumption on the operators T and L.

Assumption 2.2 (Elliptic regularity)**.** There exists a *elliptic regularity index* $\alpha \in (0, 1]$ such that

(a) T is a bounded map of $\mathbb{H}^{\alpha-1}(\Omega)$ into $\mathbb{H}^{\alpha+1}(\Omega)$,

(b) The functional F defined by

$$\langle F, v \rangle := a_\Omega(w, v), \qquad \text{for all } \ v \in H_0^1(\Omega)$$

is a bounded operator from $\mathbb{H}^{\alpha+1}$ to $\mathbb{H}^{\alpha-1}$.

When Assumption 2.2 holds for some index $\alpha \in (0, 1]$, the spaces $\mathbb{H}^r(\Omega)$ and $\dot{H}^r(\Omega)$ coincide for $r \in [-1, 1 + \alpha]$ with equivalent norms (see [23, Proposition 4.1]),

$$\|w\|_{\dot{H}^r(\Omega)} \precsim \|w\|_{\mathbb{H}^r(\Omega)} \precsim \|w\|_{\dot{H}^r(\Omega)}. \tag{2.7}$$

Moreover, inherited from the embedding property between the interpolation spaces, we have the following embedding inequality between dotted spaces:

$$\|w\|_{\dot{H}^p(\Omega)} \precsim \|w\|_{\dot{H}^q(\Omega)}, \ \text{ for } -1 \le p \le q \le 1 + \alpha. \tag{2.8}$$

2.4 Spectral fractional elliptic operators

As briefly explained in (1.1), the spectral fractional operators are defined through eigenfunction expansions. Let $\{\lambda_j, \psi_j\}$ be the eigen-pairs of L as introduced in Section 2.3. Given $s \in (0, 1)$,

we define the fractional-powered operator L^s by

$$L^s v := \sum_{j=1}^{\infty} \lambda_j^s (v, \psi_j)\psi_j \tag{2.9}$$

with domain $D(L^s) := \dot{H}^{2s}(\Omega)$. For $w, v \in \dot{H}^{2s}(\Omega)$

$$(w, v)_{\dot{H}^s} = (L^s w, v) = (w, L^s v), \quad \text{and } \|w\|_{\dot{H}^s} = (L^s w, w)^{1/2} = \left\|L^{s/2} w\right\|. \tag{2.10}$$

Of particular interest in this work is the case $s = 1/2$, as shown in the governing equations of the surface quasi-geostrophic dynamics and electroconvection system.

We now justify the extension by continuity of L^s as an operator from $\dot{H}^s(\Omega)$ to $\dot{H}^{-s}(\Omega)$. Let $w, \varphi \in \mathcal{D}(\Omega)$ where $\mathcal{D}(\Omega)$ denotes the space of distributions on Ω. From the definition (2.9) of L^s we deduce

$$\begin{aligned}(L^s w, \varphi) &= \Big(\sum_{j=1}^{\infty} \lambda_j^s b_j \psi_j,\ \sum_{k=1}^{\infty} c_k \psi_k\Big) = \sum_{j=1}^{\infty} \lambda_j^s b_j c_j \\ &= \Big(\sum_{j=1}^{\infty} \lambda_j^{s/2} b_j \psi_j,\ \sum_{k=1}^{\infty} \lambda_k^{s/2} c_k \psi_k\Big) = (L^{s/2} w, L^{s/2}\varphi) \le \|w\|_{\dot{H}^s(\Omega)} \|\varphi\|_{\dot{H}^s(\Omega)},\end{aligned} \tag{2.11}$$

where $b_j = \int_\Omega w\bar{\psi}_j$ and $c_j = \int_\Omega \varphi\bar{\psi}_j$. Hence, we define for $v, w \in \dot{H}^s(\Omega)$,

$$\langle L^s v, w\rangle := (L^{s/2} v, L^{s/2} w), \tag{2.12}$$

which extends L^s as an operator from $\dot{H}^s(\Omega)$ to $\dot{H}^{-s}(\Omega)$.

We end this section by recalling [23, Lemma 4.5] (see also [27, Lemma 3.1]) which will be used in the proof of consistency error in the finite element approximation in Chapter 3. Let $a \in [0, 1]$ and $b \in [0, 1 - a]$, then for $\mu \in (0, \infty)$

$$\left\|(\mu I + T)^{-1} v\right\|_{\dot{H}^{-2b}(\Omega)} \le \mu^{a+b-1} \|v\|_{\dot{H}^{2a}(\Omega)}, \quad \text{for all } v \in \dot{H}^{2a}(\Omega). \tag{2.13}$$

2.5 Galerkin finite element approximation

For now we assume that Ω is polyhedral. We consider a sequence of globally shape-regular, quasi-uniform, conforming subdivisions of Ω made of simplexes $\{\mathcal{T}_h : h > 0\}$. Here $h \leq 1$ denotes the maximum diameter of the subdivision $\mathcal{T}_h$. It means that for any given $h > 0$, there exists universal constants c_1 and c_2 independent of h, such that

$$\frac{\operatorname{diam}(K)}{r(K)} \leq c_1, \quad \text{for all } K \in \mathcal{T}_h, \qquad \text{and} \qquad \max_{K\in\mathcal{T}_h} \operatorname{diam}(K) \leq c_2 \min_{K\in\mathcal{T}_h} \operatorname{diam}(K), \tag{2.14}$$

where $\operatorname{diam}(K)$ stands for diameter of K and $r(K)$ for the radius of the largest ball inscribed in K. For a fixed $h > 0$, we denote V_h the space of piecewise continuous linear polynomial subordinate to $\mathcal{T}_h$ vanishing on $\partial\Omega$ and M_h the dimension of V_h.

It is in position to define the discrete counterpart of the dotted space $\dot{H}^s(\Omega)$. The numerical approximation $T_h : H^{-1}(\Omega) \to V_h$ of T is defined as its Galerkin finite element approximation on V_h: for any $F \in H^{-1}(\Omega)$, $T_hF \in V_h$ satisfies

$$a_\Omega(T_hF, \varphi_h) = \langle F, \varphi_h\rangle, \qquad \text{for all } \varphi_h \in V_h. \tag{2.15}$$

The inverse of T_h restricted to V_h is denoted by $L_h : V_h \to V_h$. We define $L_h^s : V_h \to V_h$ by

$$L_h^s v_h := \sum_{j=1}^{M_h} \lambda_{h,j}^s (v_h, \psi_{h,j})\psi_{h,j},$$

where $(\lambda_{h,j}, \psi_{h,j})_{j=1}^{M_h}$ are the eigenpairs of L_h, with $\psi_{h,j}$ $L^2(\Omega)$-orthonormal. We also define the discrete dotted norm

$$\|v_h\|_{\dot{H}_h^s(\Omega)} = (L_h^s v_h, v_h)^{1/2} = \left\| L_h^{s/2} v_h \right\|,$$

or, equivalently by

$$\|v_h\|_{\dot{H}_h^s(\Omega)} = \Big(\sum_{j=1}^{M_h} \lambda_{h,j}^s \left| (v_h, \psi_{h,j}) \right|^2 \Big)^{1/2},$$

compare with (2.6). With all the preparation, the discrete dotted space is defined by

$$\dot{H}_h^r := \dot{H}_h^r(\Omega) := \Big\{ v_h = \sum_{j=1}^{M_h} (v_h, \psi_{h,j})\psi_{h,j} \in V_h \ : \ \sum_{j=1}^{M_h} \lambda_{h,j}^r \left| (v_h, \psi_{h,j}) \right|^2 < \infty \Big\}.$$

For $r \in [0,1]$ the norms $\|\cdot\|_{\dot{H}_h^r(\Omega)}$ and $\|\cdot\|_{\dot{H}^r(\Omega)}$ are equivalent (see [40, Appendix A.2]), i.e., there exists a constant $c > 0$ independent of h, such that for all $v_h \in V_h$,

$$\frac{1}{c} \|v_h\|_{\dot{H}_h^r(\Omega)} \leq \|v_h\|_{\dot{H}^r(\Omega)} \leq c \|v_h\|_{\dot{H}_h^r(\Omega)}. \tag{2.16}$$

We next mention several inequalities that are shown to be fundamental to the argumentation provided in the rest of this dissertation. We first provide a discrete version of (2.13) (for a proof see also[27, Lemma 7.4]). Let $a \in [0,1]$ and $b \in [0, 1-a]$, then for $\mu \in (0, \infty)$

$$\left\| (\mu I + T_h)^{-1} v \right\|_{\dot{H}_h^{-2b}(\Omega)} \leq \mu^{a+b-1} \|v\|_{\dot{H}_h^{2a}(\Omega)}, \quad \text{for all } v \in \dot{H}_h^{2a}(\Omega). \tag{2.17}$$

Furthermore, we recall the estimate for $T - T_h$ provided in [23, Corollary 4.2]. For any $r_1,\, r_2 \in [0, \alpha]$ there exists a constant C independent of h, such that

$$\|(T - T_h)v\|_{\dot{H}^{1-r_1}(\Omega)} \leq C h^{r_1+r_2} \|v\|_{\dot{H}^{r_2-1}(\Omega)}, \quad \text{for all } v \in \dot{H}^{-1}(\Omega). \tag{2.18}$$

We use the notation π_h to denote the $L^2(\Omega)$-orthogonal projection onto the discrete space V_h: for any $v \in L^2(\Omega)$, $\pi_h v \in V_h$ satisfies

$$(\pi_h v, \varphi_h) = (v, \varphi_h), \qquad \text{for all } \varphi_h \in V_h. \tag{2.19}$$

Following [25, Lemma 5.1], for $r \in [0,1]$ and $\sigma \in (r, 2]$, there exists a constant $C := C(r, \sigma)$ independent of h, such that

$$\|(I - \pi_h)v\|_{\mathbb{H}^r(\Omega)} \leq C h^{\sigma - r} \|v\|_{\mathbb{H}^\sigma(\Omega)}, \qquad \text{for all } v \in \mathbb{H}^\sigma(\Omega). \tag{2.20}$$

Additionally, the quasi-uniformity assumption implies that π_h is stable from $\mathbb{H}^r(\Omega)$ onto $\mathbb{H}^r_h(\Omega)$ for $r \in [0,1]$, i.e., there exists a constant C independent of h and r, so that

$$\|\pi_h v\|_{\mathbb{H}^r_h(\Omega)} \leq C \|v\|_{\mathbb{H}^r(\Omega)}, \qquad \text{for all } v \in \mathbb{H}^r(\Omega). \tag{2.21}$$

The above inequality is still valid for $r \in (1, 1+\alpha]$, see [26, Lemma 4.1].

We also note that from the quasi-uniformity and shape-regularity assumptions on the mesh, there exists a constant C_I only dependent of mesh parameters c_1 and c_2 in (2.14), such that the following standard *inverse inequality* holds,

$$\|v_h\|_{\mathbb{H}^b(\Omega)} \leq \tilde{C}_I h^{a-b} \|v_h\|_{\mathbb{H}^a(\Omega)}, \qquad \text{for all } v_h \in V_h,\ a \leq b \leq 3/2, \tag{2.22}$$

which using the dotted space, reads

$$\|v_h\|_{\dot{H}^b(\Omega)} \leq C_I h^{a-b} \|v_h\|_{\dot{H}^a(\Omega)}, \qquad \text{for all } v_h \in \mathbb{V}_h,\ -1 \leq a \leq b \leq \min(1+\alpha, 3/2). \tag{2.23}$$

2.6 Sinc quadrature on $\mathbb{R}$

The Balakrishnan formula (1.11) suggests the necessity of seeking an efficient numerical scheme to approximate the integration over $\mathbb{R}$. The sinc quadrature scheme is chosen to achieve such purpose. Given function $f : \mathbb{R} \to \mathbb{R}$ with $I := \int_{-\infty}^{\infty} f(x)\, dx < \infty$, the sinc quadrature approximation to I is given by

$$I \approx k \sum_{j=-\infty}^{\infty} f(x_j), \tag{2.24}$$

with $k > 0$ the quadrature spacing, and $x_j = jk$. Theorem 2.20 in [24] ensures the exponential convergence property of such quadrature scheme. We summarize the results as follows.

Proposition 2.3. Define the narrow band on the complex plane $\mathcal{D}_b := \{z \in \mathbb{C} \ :\ z = t + iy,\ |y| < b\}$ with $b > 0$. Let $f : \mathbb{R} \to \mathbb{R}$ satisfy:

(a) $f(z)$ is analytic in $\mathcal{D}_b$ and continuous on the closure $\overline{\mathcal{D}_b}$,

(b) there exists an uniform constant C, such that $\int_{-b}^{b} |f(t+iy)|\ dy \le C$,

(c) $N(f, \mathcal{D}_b) := \int_{-\infty}^{\infty} |f(t+ib)| + |f(t-ib)|\, dt < \infty$.

Then the sinc quadrature error is bounded by

$$\left| \int_{-\infty}^{\infty} f(x)\, dx - k \sum_{j=-\infty}^{\infty} f(x_j) \right| \le \frac{N(f, \mathcal{D}_b)}{2 \sinh \pi b/k} e^{-2\pi b/k}. \tag{2.25}$$

If in addition the function $f(x)$ decays exponentially when $|x| \to \infty$, then we can further truncate the infinite sum in (2.24) into a finite sum, namely, by choosing two positive integers $N_- := N_-(k)$ and $N_+ := N_+(k)$ properly, we define

$$\mathcal{Q}^k f := k \sum_{j=-N_-}^{N_+} f(x_j). \tag{2.26}$$

Then the approximation error $\left| I - \mathcal{Q}^k f \right|$ converges exponentially as $k \to 0$. In the later chapters of this work we will specify the choice of parameters N_- and N_+ as well as the error estimate when we come to a specific f.

2.7 Approximation of spectral fractional elliptic operators

Following the spectral decomposition expression (2.9), for $s \in (0,1)$ and $f \in L^2(\Omega)$, the negative powered fractional operator $L^{-s} f$ is defined through replacing s by $-s$ in (2.9). In this section we briefly overview the numerical scheme to approximate $u := L^{-s} f$. We refer to [23, 25] for the detailed numerical analysis to the approximation.

The negative-powered operator $L^{-s} = T^s$ is bounded on $L^2(\Omega)$, and admits the following fundamental Dunford-Taylor representation [41]:

$$L^{-s} f = \frac{1}{2\pi i} \int_{\mathcal{C}} z^{-s} (zI - L)^{-1} f\, dz \quad \text{for } f \in L^2(\Omega). \tag{2.27}$$

with $\mathcal{C}$ a complex Jordan curve running in the resolvent set of L and oriented to have all spectrum of L to its left; see Figure 2.1(a) for example. Here $z^{-s} = e^{-s \ln z}$ with the branch cut for the

complex logarithm along the negative direction of real axis.

A proper deforming of such contour as illustrated in Figure 2.1(b) yields the following so-called Balakrishnan formula, c.f. [23],

$$u = L^{-s} f = \frac{c_s}{2} \int_0^\infty \mu^{-s} (\mu I + L)^{-1} f \, d\mu = c_s \int_{-\infty}^\infty e^{2st} (I + e^{2t} L)^{-1} f \, dt, \tag{2.28}$$

where $c_s := 2\sin(\pi s)/\pi$. Here for the second equality the change of variable $\mu = e^{-2t}$ was applied for the preparation of applying sinc quadrature approximation (2.26). Indeed, we approximate with

$$u \approx \mathcal{Q}^k u = \sum_{j=-N_-}^{N_+} e^{2st_j} w(t_j), \tag{2.29}$$

where $k > 0$ is the sinc quadrature stepping, $t_j = jk$,

$$N_+ := \left\lceil \frac{\pi^2}{sk^2} \right\rceil \text{ and } N_- := \left\lceil \frac{\pi^2}{(1-s)k^2} \right\rceil. \tag{2.30}$$

Here for each j the function $w(t_j) = w(t_j; f) := (I + e^{2t_j} L)^{-1} f$ solves

$$(w, v)_\Omega + e^{2t_j} a_\Omega(w, v) = (f, v)_\Omega, \qquad \text{for all } v \in H_0^1(\Omega), \tag{2.31}$$

with the bilinear form $a_\Omega(\cdot, \cdot)$ associated to the definition of L as in Section 2.3.

The solutions $w(t_j)$, $j = -N_-, \cdots, N_+$ of the subproblems (2.31) are in turn approximated by continuous piecewise linear functions. Suppose that Ω admits a sequence of subdivisions $\{\mathcal{T}_h : h > 0\}$ as proposed in Subsection 2.5. For a fixed $h > 0$, let V_h be the space of piecewise linear continuous polynomials subordinate to $\mathcal{T}_h$ vanishing on $\partial\Omega$. We are in position to propose the finite element approximation u_h of u by

$$u_h := \sum_{j=-N_-}^{N_+} e^{2st_j} w_h(t_j), \tag{2.32}$$

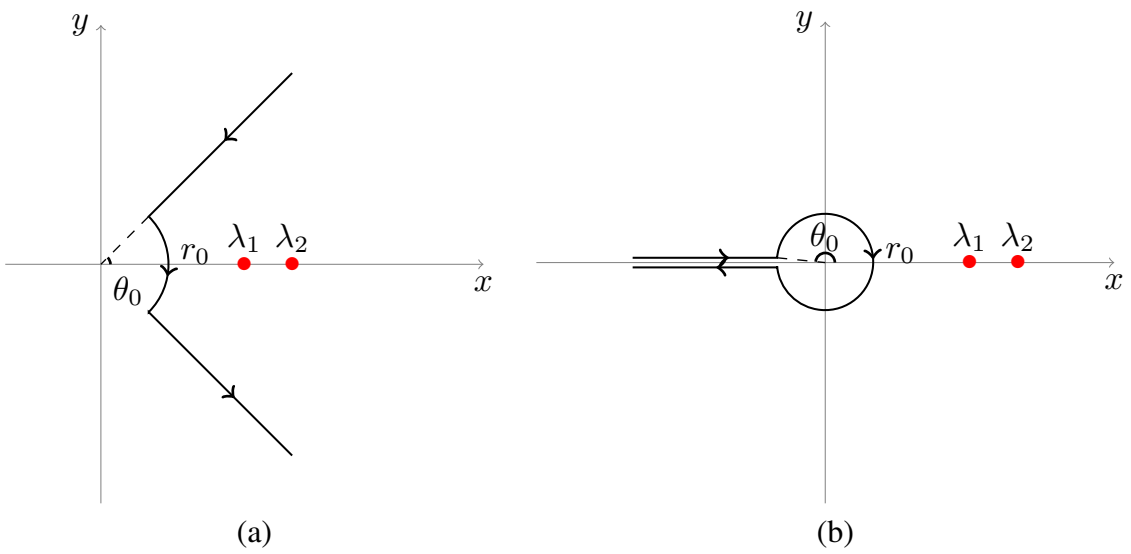


Figure 2.1. The contours of $\mathcal{C}$. (a) Contour used in the Dunford-Taylor representation, and (b) Contour used in the Balakrishnan representation. Contour (b) is obtained by letting $\theta_0 \to \pi$ and $r_0 \to 0$ in Contour (a).

where $w_h(t_j) := w_h(t_j; f) \in V_h$ solves

$$(w_h, \varphi_h) + e^{2t_j} a_\Omega(w_h, \varphi_h) = (f, \varphi_h), \qquad \text{for all } \varphi_h \in V_h. \tag{2.33}$$

It is worth mentioning that the $N_- + N_+ + 1$ finite element problems (2.33) are mutually *independent*, making the parallel implementation straightforward. In particular, for each subproblem (2.33) it only requires the implementation of a classical finite element solver for diffusion-reaction problems. Also, the algorithm consists of an outer loop (2.32) gathering the contributions of the finite element solutions at each quadrature point t_j.

The error between u in (2.28) and its finite element approximation u_h defined by (2.32) consists of the exponentially-convergent sinc quadrature error and the polynomial-order finite element approximation error.

Theorem 2.4 (Theorems 3.5 and 4.3 of [23])**.** Let Assumption 2.2 hold for index $\alpha \in (0, 1]$. Set $\gamma = \alpha - s$ when $\gamma \geq s$, and $\gamma = 0$ when $\gamma < s$. Let k be the sinc quadrature stepping such that N_- and N_+ are defined by (2.30). For $f \in \mathbb{H}^{2\delta}(\Omega)$ with $\delta \geq \gamma$, there exists a constant C independent

of k, δ, and h satisfying

$$\|u - u_h\| \leq C_h \left(e^{-\pi^2/(2k)} + h^{2\alpha} \right) \|f\|_{\mathbb{H}^{2\delta}(\Omega)}. \tag{2.34}$$

Here

$$C_h = \begin{cases} C \ln(1/h), & \text{if } \delta = \gamma \text{ and } \alpha \geq s, \\ C, & \text{if } \delta > \gamma \text{ and } \alpha \geq s, \\ C, & \text{if } s > \alpha. \end{cases}$$

2.8 Approximation of integral fractional Laplacians

Unlike the spectral fractional Laplacian, there is no direct integral representation for the solution u to the problem (1.6) with $L^s = (-\Delta)^s$ the integral fractional laplacian. Instead, the equation could be solved iteratively by implementing the action of the operator $(-\Delta)^s \tilde{u} \in \mathbb{H}^{-s}(\Omega)$ on the space $\mathbb{H}^s(\Omega)$. It turns out that the action can be represented with the following Balakrishnan formula (cf. [27, Theorem 4.1], also compare with (2.28))

$$a_{\mathscr{F}}(u, \varphi) := ((-\Delta)^s \tilde{u}, (-\Delta)^s \tilde{\varphi})_{\mathbb{R}^d} = \int_{-\infty}^{\infty} e^{st} (w(t),\, \varphi)_\Omega \, dt, \tag{2.35}$$

where for any $t \in \mathbb{R}$, the function $w(t) := (-\Delta)(e^t I - \Delta)^{-1} \tilde{u} = v(t) + u$ with $v(t) \in H^1(\mathbb{R}^d)$ solves

$$e^t (v(t), \varphi)_{\mathbb{R}^d} + (\nabla v(t), \nabla \varphi)_{\mathbb{R}^d} = -e^t (u, \varphi)_\Omega, \qquad \text{for all } \varphi \in H^1(\mathbb{R}^d). \tag{2.36}$$

Similar to the the approximation of the spectral fractional Laplacian case, we can now use a sinc quadrature and standard finite element methods to provide an approximation of (2.35). The additional caveat steams from the fact that the independent problems (2.36) need to be solved in the whole plane $\mathbb{R}^2$. To circumvent this issue, we replace (2.36) with another problem imposed on the bounded domain with vanishing boundary condition. Observe that since $\text{supp}(u) \subset \Omega$, $w(t)$ decays to zero exponentially as $|x| \to \infty$, thus $w(t)$ can be approximated by $w^M(t) := u + v^M(t)$

with $v^M(t) \in H_0^1(\Omega^M(t))$ solves

$$e^t(v^M(t),\varphi)_{\Omega^M(t)} + (\nabla v^M(t), \nabla\varphi)_{\Omega^M(t)} = -e^t(u,\varphi)_\Omega, \qquad \text{for all } \varphi \in H_0^1(\Omega^M(t)).$$

Here the truncated domain $\Omega^M(t)$ is defined by

$$\Omega^M(t_j) := \begin{cases} \left\{1 + e^{-t/2}(M+1)x \ : \ x \in \Omega\right\}, & e^{-t/2} \geq 1, \\ \left\{(M+2)x \ : \ x \in \Omega\right\}, & e^{-t/2} < 1. \end{cases} \tag{2.37}$$

Suppose that u in (2.35) satisfies $u \in \mathbb{H}^\delta(\Omega)$ with $\delta \in (s, 2-s]$. After applying the sinc quadrature, the infinite integral is approximated by

$$a_{\mathscr{F}}(u,\varphi) \approx \tilde{a}^{k,M}(u,\varphi) := \frac{c_s}{2} \sum_{j=-N_-}^{N_+} e^{st_j}(w^M(t_j),\, \varphi)_\Omega,$$

where for the given quadrature spacing $k > 0$,

$$N_+ = \left\lceil \frac{\pi^2}{(\delta - s)k^2} \right\rceil, \quad \text{and} \quad N_- = \left\lceil \frac{\pi^2}{2sk^2} \right\rceil. \tag{2.38}$$

The last step of approximation consists of the application of finite element method. For *each* of the truncated domain $\Omega^M(t_j)$, we discretize it into a sequence of globally shape-regular, quasi-uniform, conforming subdivisions made of simplexes $\{\mathcal{T}_h^M(t_j) \ : \ h > 0\}$. For a fixed $h > 0$ the finite element space $V_h^M(t_j)$ associated to $\Omega^M(t_j)$ is therefore defined to be the space of continuous piece-wise linear polynomials vanishing on the boundary of $\Omega^M(t_j)$. For $u_h, \varphi_h \in V_h(\Omega)$, the finite element approximation of $\tilde{a}^{k,M}(u,\varphi)$ is defined by

$$a_{\mathscr{F}}^{k,M,h}(u_h,\varphi_h) = \frac{c_s}{2} \sum_{j=-N_-}^{N_+} e^{st_j} \left(w_h^M(t_j), \varphi_h\right)_\Omega, \tag{2.39}$$

with $w_h^M(t_j) = \tilde{u}_{h|\Omega^M(t_j)} + v_h^M(t_j)$ for each j, and $v_h^M(t_j)$ solves

$$e^{t_j}(v_h^M(t_j), \phi_h)_{\Omega^M(t_j)} + (\nabla v_h^M(t_j), \nabla \phi_h)_{\Omega^M(t_j)} = -e^{t_j}(\tilde{u}_h, \phi_h)_{\Omega^M(t_j)}, \text{ for all } \phi_h \in V_h^M(t_j).$$

In view of (1.6), assume that $u : \Omega \to \mathbb{R}$ solves

$$a_{\mathscr{F}}(u, \varphi) = (f, \varphi) \qquad \text{for all } \varphi \in \mathbb{H}^s(\Omega), \tag{2.40}$$

and $u_h : \Omega \to \mathbb{R}$ solves

$$a_{\mathscr{F}}^{k,M,h}(u_h, \varphi_h) = (f, \varphi_h) \qquad \text{for all } \varphi_h \in V_h(\Omega),$$

then the L^2 error between u and u_h consists of the errors from the sinc quadrature stepping, the truncated domain problem approximation, and the finite element approximation.

Theorem 2.5 (Theorem V.14 of [42])**.** Assume $\partial\Omega$ is of C^∞ class. Assume that solution u of (1.6) belongs to $\mathbb{H}^\beta(\Omega)$ with $\beta \in (s, 3/2)$. Let $\delta = \min(2 - s, \beta)$, and k be the sinc quadrature stepping such that N_- and N_+ are defined as in (2.38). Assume k is chosen sufficiently small to ensure the $V_h(\Omega)$ ellipticity of $a_{\mathscr{F}}^{k,M,h}(\cdot, \cdot)$. Then there exist constants c and C independent of M, k, and h, such that

$$\begin{aligned} \|u - u_h\| \leq C \ln(h^{-1}) \Big(&(\frac{1}{\delta - s} + \frac{1}{s}) e^{-\pi^2/(2k)} + e^{-cM} \\ &+ \ln(h^{-1}) h^{\beta + \min(0, 1/2 - s) - \epsilon} \Big) \|u\|_{\mathbb{H}^\beta(\Omega)}, \end{aligned} \tag{2.41}$$

with $\epsilon > 0$.

3. APPROXIMATION OF THE STATIONARY PROBLEM

In this chapter, we study the stationary problem: find $w : \Omega \to \mathbb{R}$ satisfying

$$\boldsymbol{u} \cdot \nabla w + L^s w = f \quad \text{in } \Omega, \qquad w = 0 \quad \text{on } \partial\Omega. \tag{3.1}$$

where $s \in [1/2, 1)$, $f \in L^2(\Omega)$, and $\boldsymbol{u} \in \boldsymbol{U}$ is given with $\boldsymbol{U}$ defined by (1.5). The fractional operator L^s could be either spectral fractional operator, or integral fractional Laplacian. The results obtained in the stationary case will be instrumental to the derivation of error estimates for the time dependent problem in Chapter 4.

The content of this chapter is outlined as follows. In Section 3.1 we provide a weak formulation to (3.1), together with a regularity discussion on the weak solution. An integral representation followed by a sinc quadrature approximation is given in Section 3.2. The sinc quadrature error is provided therein. Section 3.3 discusses the finite element approximation and the corresponding consistency error. The energy and L^2 errors for the approximation are presented in Section 3.4. Numerical results are reported in Section 3.5.

3.1 Weak formulation

We have already established that $\langle L^s w, \varphi \rangle = (L^{s/2} w, L^{s/2} \varphi)$ for $w,\ \varphi \in \dot{H}^s$, see (2.12). Hence we now focus on the well-poseness of the nonlinear term $\boldsymbol{u} \cdot \nabla w$. As $\boldsymbol{u} \in \boldsymbol{U}$, we observe that for $w, \varphi \in \mathcal{D}(\Omega)$, the space of distributions on Ω,

$$(\boldsymbol{u} \cdot \nabla w, \varphi) = -(\boldsymbol{u} \cdot \nabla \varphi, w). \tag{3.2}$$

On the one hand, this implies that

$$\|\boldsymbol{u} \cdot \nabla w\|_{H^{-1}} = \sup_{0 \neq \psi \in \mathcal{D}(\Omega)} \frac{(\boldsymbol{u} \cdot \nabla w, \psi)}{\|\psi\|_{H^1}} = \sup_{0 \neq \psi \in \mathcal{D}(\Omega)} \frac{-(\boldsymbol{u} \cdot \nabla \psi, w)}{\|\psi\|_{H^1}} \leq \|\boldsymbol{u}\|_{\mathbf{L}^\infty} \|w\|,$$

while on the other hand, denoting by u_j, $j = 1, \cdots, d$ the components of $\boldsymbol{u}$, we have

$$\|\boldsymbol{u} \cdot \nabla w\| = \left\| \sum_{j=1}^{d} u_j \partial_j w \right\| \leq \|\boldsymbol{u}\|_{\mathbf{L}^\infty} \|\nabla w\| \leq \|\boldsymbol{u}\|_{\mathbf{L}^\infty} \|w\|_{H^1} .$$

Hence, Proposition 2.1 guarantees that

$$\|\boldsymbol{u} \cdot \nabla w\|_{\mathbb{H}^{-r}} \leq \|\boldsymbol{u}\|_{\mathbf{L}^\infty} \|w\|_{\mathbb{H}^{1-r}} , \tag{3.3}$$

for any $r \in [0, 1]$. Because the dotted spaces $\dot{H}^r$ and Sobolev spaces $\mathbb{H}^r$ have equivalent norms for $-1 \leq r \leq 1$, (3.3) can be rewritten as

$$\|\boldsymbol{u} \cdot \nabla w\|_{\dot{H}^{-r}} \precsim \|\boldsymbol{u}\|_{\mathbf{L}^\infty} \|w\|_{\dot{H}^{1-r}} . \tag{3.4}$$

As $\mathcal{D}(\Omega)$ is dense in the Sobolev space $\mathbb{H}^r$, or in the dotted space $\dot{H}^r$ for $0 \leq r \leq 1$, the mapping $w \mapsto \boldsymbol{u} \cdot \nabla w$ is well-defined as an operator from $\dot{H}^s$ to $\dot{H}^{-s}$, and

$$\|\boldsymbol{u} \cdot \nabla w\|_{\dot{H}^{-s}} \precsim \|\boldsymbol{u}\|_{\mathbf{L}^\infty} \|w\|_{\dot{H}^{1-s}} \precsim \|\boldsymbol{u}\|_{\mathbf{L}^\infty} \|w\|_{\dot{H}^s} , \tag{3.5}$$

where we used the embedding $\dot{H}^s \subset \dot{H}^{1-s}$ when $s \geq 1/2$.

We are now in position to write a weak formulation of Equation (3.1): for $s \in [1/2, 1)$ and given $\boldsymbol{u} \in \mathbf{U}$, find $w \in V := \dot{H}^s$ such that

$$a(w, \varphi) := \langle \boldsymbol{u} \cdot \nabla w, \varphi \rangle + (L^{s/2} w, L^{s/2} \varphi) = (f, \varphi), \qquad \text{for all } \varphi \in V. \tag{3.6}$$

To derive the coercivity of $a(\cdot, \cdot)$, it suffices to note that for $\boldsymbol{u} \in \boldsymbol{U}$

$$\langle \boldsymbol{u} \cdot \nabla w, \varphi \rangle = \frac{1}{2} \langle \boldsymbol{u} \cdot \nabla w, \varphi \rangle - \frac{1}{2} \langle \boldsymbol{u} \cdot \nabla \varphi, w \rangle, \ \text{ for all } \varphi \in \dot{H}^s, \tag{3.7}$$

there holds $\langle \boldsymbol{u} \cdot \nabla w, w \rangle = 0$, and therefore $a(w, w) = (L^{s/2} w, L^{s/2} w) = \|w\|^2_{\dot{H}^s}$. The boundedness

of $a(\cdot,\cdot)$ follows from (2.10) and (3.5). Indeed,

$$a(w,\varphi) \le |\langle \boldsymbol{u}\cdot\nabla w,\varphi\rangle| + \|w\|_{\dot H^s}\|\varphi\|_{\dot H^s} \precsim \|\boldsymbol{u}\|_{\mathbf{L}^\infty}\|w\|_{\dot H^{1-s}}\|\varphi\|_{\dot H^s} + \|w\|_{\dot H^s}\|\varphi\|_{\dot H^s}$$
$$\precsim (\|\boldsymbol{u}\|_{\mathbf{L}^\infty}+1)\|w\|_{\dot H^s}\|\varphi\|_{\dot H^s}\,.$$

Lax-Milgram theory ensures the existence and uniqueness of the solution to (3.6) in $\dot H^s$.

The next proposition shows that the solution $w\in\dot H^s$ of (3.6) is, in fact, more regular.

Proposition 3.1 (Improved regularity)**.** Let $s\in[1/2,1)$, $f\in L^2(\Omega)$, and assume w solves the weak problem (3.6) with $\boldsymbol{u}\in\mathbf{U}$. Then there exist $\rho\in(0,1/2)$ and a constant $C_{\max}$ both independent of $\boldsymbol{u}$, such that $w\in\dot H^{2s^*}$ and $\|w\|_{\dot H^{2s^*}}\precsim\|f\|$, with

$$s^* = \begin{cases} s, & \text{if } s\in(1/2,1),\\ 1/2, & \text{if } s=1/2 \text{ and } \|\boldsymbol{u}\|_{\mathbf{L}^\infty}\le C_{\max},\\ (1-\rho)/2, & \text{if } s=1/2 \text{ and } \|\boldsymbol{u}\|_{\mathbf{L}^\infty}> C_{\max}. \end{cases} \tag{3.8}$$

Proof. We distinguish two cases depending on the value of s.

1 We start with the case $s\in(1/2,1)$. The weak solution $w\in\dot H^s$ of (3.6) satisfies $\|w\|_{\dot H^s}\le\|f\|$ and

$$(L^{s/2}w, L^{s/2}\varphi) = \langle L^s w,\varphi\rangle = F(\varphi),\ \text{ for all } \varphi\in\dot H^s, \tag{3.9}$$

where, in view of (3.5), the functional $F\in\dot H^{-s}$ defined by $F(\varphi):=-\langle\boldsymbol{u}\cdot\nabla w,\varphi\rangle+\langle f,\varphi\rangle$ belongs to $\dot H^{-1+s}$. Whence the lifting property of L^s on the dotted space implies $w\in\dot H^{-1+3s}$ and

$$\|w\|_{\dot H^{-1+3s}}\precsim(1+\|\boldsymbol{u}\|_{\mathbf{L}^\infty})\|f\|\,.$$

In general, if $w\in\dot H^\gamma$ and $\|w\|_{\dot H^\gamma}\precsim\|f\|$ for $\gamma\in(s,2s)$, letting $r:=1-\min(\gamma,1)$ in (3.4), we have $\|\boldsymbol{u}\cdot\nabla w\|_{\dot H^{\min(\gamma-1,0)}}\precsim\|\boldsymbol{u}\|_{\mathbf{L}^\infty}\|w\|_{\dot H^{\min(\gamma,1)}}<\infty$, which yields $w\in\dot H^{\min(\gamma+2s-1,2s)}$ and

$$\|w\|_{\dot H^{\min(\gamma+2s-1,2s)}}\precsim\|f\|+\|\boldsymbol{u}\|_{\mathbf{L}^\infty}\|w\|_{\dot H^{\min(\gamma,1)}}\precsim\|f\|\,.$$

Proceeding by induction leads to $w \in \dot{H}^{2s}$ and $\|w\|_{\dot{H}^{2s}} \precsim \|f\|$.

2 When $s = 1/2$, we use the notation $\Lambda := L^{1/2}$ and rewrite (3.6) as

$$\langle (I - B)w, \varphi \rangle = k \langle \Lambda^{-1} f, \varphi \rangle, \quad \text{for all } \varphi \in \dot{H}^s, \tag{3.10}$$

where $B := I - k\Lambda^{-1}(\Lambda + \boldsymbol{u} \cdot \nabla)$ with $k \in (0, 1)$ a constant dependent of $\|\boldsymbol{u}\|_{\mathbf{L}^\infty}$ yet to be determined. From the definition of dotted inner product (2.10) and the interpolation inequality (3.4) we deduce that

$$\left\|\Lambda^{-1}(\boldsymbol{u} \cdot \nabla)\theta\right\|_{\dot{H}^\gamma} = \|(\boldsymbol{u} \cdot \nabla)\theta\|_{\dot{H}^{\gamma-1}} \leq C \|\boldsymbol{u}\|_{\mathbf{L}^\infty} \|\theta\|_{\dot{H}^\gamma}, \qquad \text{for } \gamma \in [0, 1]$$

with C independent of $\boldsymbol{u}$, θ and γ. It additionally follows from $(\boldsymbol{u} \cdot \nabla w, w) = 0$ and (2.10) that

$$(\theta, \Lambda^{-1}(\boldsymbol{u} \cdot \nabla)\theta)_{\dot{H}^{1/2}} = (\theta, (\boldsymbol{u} \cdot \nabla)\theta) = 0, \qquad \text{and} \qquad (\Lambda^{-1}(\boldsymbol{u} \cdot \nabla)\theta, \theta)_{\dot{H}^{1/2}} = ((\boldsymbol{u} \cdot \nabla)\theta, \theta) = 0.$$

Hence for any $\theta \in \dot{H}^{1/2}$,

$$\begin{aligned} \|B\theta\|_{\dot{H}^{1/2}}^2 &= \|\theta\|_{\dot{H}^{1/2}}^2 - k(\theta, \Lambda^{-1}(\Lambda + \boldsymbol{u} \cdot \nabla)\theta)_{\dot{H}^{1/2}} \\ &\quad - k(\Lambda^{-1}(\Lambda + \boldsymbol{u} \cdot \nabla)\theta, \theta)_{\dot{H}^{1/2}} + k^2 \left\|\Lambda^{-1}(\Lambda + \boldsymbol{u} \cdot \nabla)\theta\right\|_{\dot{H}^{1/2}}^2 \\ &\leq \left[1 - 2k + (C \|\boldsymbol{u}\|_{\mathbf{L}^\infty} + 1)^2 k^2\right] \|\theta\|_{\dot{H}^{1/2}}^2. \end{aligned} \tag{3.11}$$

The coefficient $1 - 2k + (C \|\boldsymbol{u}\|_{\mathbf{L}^\infty} + 1)^2 k^2$ attains its minimum

$$M_1 := 1 - (C \|\boldsymbol{u}\|_{\mathbf{L}^\infty} + 1)^{-2} < 1$$

when $k = (C \|\boldsymbol{u}\|_{\mathbf{L}^\infty} + 1)^{-2}$. In addition, for any $\theta \in \dot{H}^1$

$$\|B\theta\|_{\dot{H}^1} = \left\|(1 - k)\theta - k\Lambda^{-1}\boldsymbol{u} \cdot \nabla\theta\right\|_{\dot{H}^1} \leq [1 + (C \|\boldsymbol{u}\|_{\mathbf{L}^\infty} - 1)k] \|\theta\|_{\dot{H}^1} =: M_2 \|\theta\|_{\dot{H}^1}, \tag{3.12}$$

We set $C_{\max} = 1/C$ so that if $\|\boldsymbol{u}\|_{\mathbf{L}^\infty} < C_{\max}$, then $M_2 < 1$ and $I - B$ is invertible in $\dot{H}^1$. We find that $w \in \dot{H}^1$ and $\|w\|_{\dot{H}^1} \precsim \|f\|$ from (3.10). When instead $\|\boldsymbol{u}\|_{\mathbf{L}^\infty} \geq C_{\max}$, we invoke the operator interpolation theory to deduce from (3.11) and (3.12) that

$$\|B\|_{\dot{H}^{1-\rho} \to \dot{H}^{1-\rho}} \leq M_1^{2\rho} M_2^{1-2\rho}, \text{ for all } \rho \in [0, 1/2].$$

Restricting $\rho > \rho_0$ with

$$\rho_0 := \frac{1}{2} \log \frac{(C\|\boldsymbol{u}\|_{\mathbf{L}^\infty} + 1)^2}{(C\|\boldsymbol{u}\|_{\mathbf{L}^\infty} + 1)^2 - (C\|\boldsymbol{u}\|_{\mathbf{L}^\infty} - 1)} \left[\log \frac{(C\|\boldsymbol{u}\|_{\mathbf{L}^\infty} + 1)^2 - 1}{(C\|\boldsymbol{u}\|_{\mathbf{L}^\infty} + 1)^2 - (C\|\boldsymbol{u}\|_{\mathbf{L}^\infty} - 1)} \right]^{-1},$$

we realize that $M_1^{2\rho} M_2^{1-2\rho} < 1$, and so $I - B$ is invertible in $\dot{H}^{1-\rho}$. Returning to (3.10), this infers that $w \in \dot{H}^{1-\rho}$ and $\|w\|_{\dot{H}^{1-\rho}} \precsim \|f\|$. □

With the regularity pickup in Assumption 2.2, the proposition can be translated into standard Sobolev spaces. This is the subject of the following corollary.

Corollary 3.2. Let $s \in [1/2, 1)$, $f \in L^2(\Omega)$, and assume w solves (3.1) with $\boldsymbol{u} \in \boldsymbol{U}$. When Assumption 2.2 holds with parameter $\alpha \in (0, 1]$, for the same constants ρ, $C_{\max}$ and s^* as in Proposition 3.1, we have $w \in \mathbb{H}^{\min(2s^*, 1+\alpha)}$.

Proof. This follows from the norm equivalence (2.7). □

Remark 3.3 (Periodic boundary condition)**.** For the periodic boundary problem imposed on $\Omega = (0, 1)^d$ and $L = -\Delta$, there will be no loss of regularity. Assume $s \in [1/2, 1)$, $f \in L^2(\Omega)$, $\boldsymbol{u} \in \boldsymbol{U}$, and assume w solves (3.6) with w periodic. Then $w \in \mathbb{H}_p^{2s}(\Omega)$.

Proof. Exploiting the Fourier series expansions we have

$$w = \sum_{|j| \neq 0} \hat{w}(j) e^{2\pi i x \cdot j}, \qquad \text{and } f = \sum_{|j| \neq 0} \hat{f}(j) e^{2\pi i x \cdot j},$$

where $\hat{w}(j) = \int_\Omega e^{-2\pi i x\cdot j} w(x)\,dx$ and $\hat{f}(j) = \int_\Omega e^{-2\pi i x\cdot j} f(x)\,dx$, then

$$\boldsymbol{u}\cdot\nabla w + (-\Delta)^s w = \sum_{|j|\neq 0} \hat{w}(j)\left(2\pi(j\cdot\boldsymbol{u})i + (2\pi\,|j|)^{2s}\right) e^{2\pi i x\cdot j}.$$

A comparison between terms in the Fourier expansions to $\boldsymbol{u}\cdot\nabla w + (-\Delta)^s w$ and f yields

$$\hat{w}(j) = \frac{1}{2\pi(j\cdot\boldsymbol{u})i + (2\pi\,|j|)^{2s}}\hat{f}(j) = \frac{-2\pi(j\cdot\boldsymbol{u})i + (2\pi\,|j|)^{2s}}{(2\pi j\cdot\boldsymbol{u})^2 + (2\pi\,|j|)^{4s}}\hat{f}(j).$$

It follows that

$$\|w\|^2_{\dot{H}^{2s}_p(\Omega)} = \sum_{|j|\neq 0} \frac{(-2\pi(j\cdot\boldsymbol{u})i + (2\pi\,|j|)^{2s})^2(2\pi\,|j|)^{4s}}{((2\pi j\cdot\boldsymbol{u}))^2 + (2\pi\,|j|)^{4s})^2}\left|\hat{f}(j)\right|^2 \precsim \|f\|^2_{L^2(\Omega)}.$$

The conclusion follows from the norm equivalence between $\|\cdot\|_{\dot{H}^{2s}_p}$ and $\|\cdot\|_{\mathbb{H}^{2s}_p}$ upon noting that the elliptic regularity index $\alpha = 1$. □

3.2 Integral representation and sinc quadrature

In this section we will derive an integral representation for the term $(L^{s/2}w, L^{s/2}\varphi)$ in the bilinear form $a(\cdot,\cdot)$ defined in (3.6), which will be further approximated in a second step using the sinc quadrature introduced in Section 2.6.

Lemma 3.4 (Integral representation)**.** Let $0 \le r \le s < 1$. For $w \in \dot{H}^{s+r}$ and $\varphi \in \dot{H}^{s-r}$, we have

$$\left(L^{(s+r)/2}w, L^{(s-r)/2}\varphi\right) = c_s \int_0^\infty \mu^{1-2s}\left(L(I+\mu^2 L)^{-1}w, \varphi\right)\,d\mu, \tag{3.13}$$

with c_s defined by

$$c_s := \left(\int_0^\infty \mu^{1-2s}(1+\mu^2)^{-1}\,d\mu\right)^{-1} = \frac{2\sin(\pi s)}{\pi}. \tag{3.14}$$

Proof. The proof follows the argumentation in [27, Theorem 4.1] but using Fourier series instead

of using Fourier transforms. We write w and φ in the $L^2(\Omega)$ orthonormal basis $\{\psi_i\}_{i=1}^{\infty}$ and obtain

$$\big(L(I+\mu^2 L)^{-1}w,\varphi\big) = \left(\sum_{j=1}^{\infty}\frac{\lambda_j}{1+\mu^2\lambda_j}b_j\psi_j,\ \sum_{k=1}^{\infty}d_k\psi_k\right) = \sum_{j=1}^{\infty}\frac{\lambda_j}{1+\mu^2\lambda_j}b_jd_j, \tag{3.15}$$

with coefficients $b_j := \int_\Omega w\psi_j$ and $d_j := \int_\Omega \varphi\psi_j$. Recalling that $\lambda_j > 0$, therefore by Tonelli's lemma [43],

$$\begin{aligned} A := \int_0^\infty \sum_{j=1}^{\infty}\left|\frac{\lambda_j\mu^{1-2s}}{1+\mu^2\lambda_j}b_jd_j\right|\,d\mu &= \int_0^\infty\sum_{j=1}^{\infty}\frac{\lambda_j\mu^{1-2s}}{1+\mu^2\lambda_j}\,|b_j|\,|d_j|\,d\mu \\ &= \sum_{j=1}^{\infty}|b_j|\,|d_j|\int_0^\infty\frac{\lambda_j\mu^{1-2s}}{1+\mu^2\lambda_j}\,d\mu. \end{aligned} \tag{3.16}$$

Introducing the change of variable $\xi = \lambda_j^{1/2}\mu$ and using the definition (3.14) of c_s we obtain

$$\int_0^\infty\frac{\lambda_j\mu^{1-2s}}{1+\mu^2\lambda_j}\,d\mu = \lambda_j^s\int_0^\infty\frac{\xi^{1-2s}}{1+\xi^2}\,d\xi = \lambda_j^s c_s^{-1}. \tag{3.17}$$

Consequently, inserting the above equality back into (3.16) we have

$$\begin{aligned} A &= c_s^{-1}\sum_{j=1}^{\infty}|b_j|\,|d_j|\,\lambda_j^s \le c_s^{-1}\Big(\sum_{j=1}^{\infty}|b_j|^2\,\lambda_j^{(s+r)/2}\Big)^{1/2}\Big(\sum_{j=1}^{\infty}|d_j|^2\,\lambda_j^{(s-r)/2}\Big)^{1/2} \\ &= c_s^{-1}\,\|w\|_{\dot H^{s+r}}\,\|\varphi\|_{\dot H^{s-r}} < \infty, \end{aligned}$$

which ensures that the integral representation in (3.13) is well-defined. We are now eligible to use Fubini's lemma [43] to conclude that

$$\begin{aligned} c_s\int_0^\infty\mu^{1-2s}\,\big(L(I+\mu^2L)^{-1}w,\varphi\big)\,d\mu &= c_s\int_0^\infty\sum_{j=1}^{\infty}\frac{\lambda_j\mu^{1-2s}}{1+\mu^2\lambda_j}b_jd_j\,d\mu \\ = c_s\sum_{j=1}^{\infty}b_jd_j\int_0^\infty\frac{\lambda_j\mu^{1-2s}}{1+\mu^2\lambda_j}\,d\mu &= \sum_{j=1}^{\infty}\lambda_j^{(s+r)/2}b_j\lambda_j^{(s-r)/2}d_j = \big(L^{(s+r)/2}w, L^{(s-r)/2}\varphi\big), \end{aligned}$$

where in the third equality we used (3.17). This completes the proof. □

The above lemma ensures that for any $w, \varphi \in \dot{H}^s$,

$$\begin{aligned}\left(L^{s/2}w, L^{s/2}\varphi\right) &= c_s \int_0^\infty \mu^{1-2s} \left(L(I+\mu^2 L)^{-1}w, \varphi\right)\, d\mu \\ &= \frac{c_s}{2} \int_{-\infty}^\infty e^{(1-s)t} \left(L(I+e^t L)^{-1}w, \varphi\right)\, dt,\end{aligned} \tag{3.18}$$

where we use the change of variable $\mu = e^{t/2}$ to derive the second equality for the preparation of applying sinc quadrature approximation.

Following Section 2.6, (see also, e.g. [23, 25, 29, 27]), we now apply a sinc quadrature scheme to approximate the infinite integral with respect to t in (3.18):

$$\left(L^{s/2}w, L^{s/2}\varphi\right) \approx (\mathcal{Q}_k^s w, \varphi) := \frac{c_s}{2} k \sum_{j=-N_-}^{N_+} e^{(1-s)t_j} \left(L(I+e^{t_j}L)^{-1}w, \varphi\right), \tag{3.19}$$

where k is the quadrature step size, $t_j = jk$, N_- and N_+ are integer parameters specified later (see Remark 3.6). This leads to the following approximation of the bilinear form

$$a(w, \varphi) \approx a^k(w, \varphi) = \langle \boldsymbol{u} \cdot \nabla w, \varphi \rangle + (\mathcal{Q}_k^s w, \varphi). \tag{3.20}$$

Such approximation entails a consistency error we analyze now, again following Section 2.6.

Lemma 3.5 (sinc quadrature consistency error)**.** Suppose $a(w, \varphi)$ and $a^k(w, \varphi)$ are bilinear forms defined by (3.6) and (3.20) respectively, let $\varepsilon \ll 1$ be an arbitrary positive constant, and suppose $w \in \dot{H}^\delta$, $\varphi \in \dot{H}^s$ with $\delta \in (s, 2-s]$, then

$$\begin{aligned}\left|(a - a^k)(w, \varphi)\right| \leq & \frac{K_\varepsilon}{s} \left(\frac{e^{-\pi^2/k + \pi\varepsilon/k}}{\sinh(\pi^2/k - \pi\varepsilon/k)} + e^{-N_+ ks} \right) \|w\|_{L^2} \|\varphi\|_{L^2} \\ & + \frac{2K_\varepsilon}{\delta - s} \left(\frac{e^{-\pi^2/k + \pi\varepsilon/k}}{\sinh(\pi^2/k - \pi\varepsilon/k)} + e^{-N_+ k(\delta - s)/2} \right) \|w\|_{\dot{H}^\delta} \|\varphi\|_{\dot{H}^s},\end{aligned} \tag{3.21}$$

with $K_\varepsilon := (\frac{2}{1-\cos\epsilon})^{1/2}$.

Proof. We denote the integrand

$$\Psi(t; w, \varphi) := e^{(1-s)t} \left(L(I + e^t L)^{-1} w, \varphi \right),$$

and in order to apply Proposition 2.3 we are required to justify the three conditions. In particular we choose $b = \pi - \varepsilon$ for the band width of $\mathcal{D}_b$.

The proof of condition (a) is already contained in the proof in [27, Theorem 5.1] using the analyticity of operator mapping $z \to (I + e^z L)^{-1}$ on $\mathcal{D}_b$. We now prove conditions (b) and (c). We first observe that for any $z = t + iy \in \mathcal{D}_b$ and any $\lambda > 0$, if $|y| \leq \pi/4$, we have $|e^{-z} + \lambda| \geq |e^{-t} \cos y + \lambda| \geq \cos(\pi/4)(e^{-t} + \lambda)$. Whereas for $|y| \in (\pi/4, \pi - \epsilon]$,

$$\left| e^{-z} + \lambda \right|^2 = \lambda^2 + e^{2t} + 2\lambda e^t \cos y \geq (\lambda^2 + e^{2t})(1 - |\cos y|) \geq \frac{1 - \cos \epsilon}{2} (e^{-t} + \lambda)^2.$$

Combining them together, we obtain $|e^{-z} + \lambda| \geq K_\varepsilon^{-1}(e^{-t} + \lambda)$ for any $z \in \mathcal{D}_b$. With this lower bound and the expansion of w, φ in the basis $\{\psi_i\}_{i=1}^\infty$, we deduce that for any $t \in \mathbb{R}$ and $|y| \leq \pi - \varepsilon$

$$|\Psi(t + iy; w, \varphi)| \leq K_\varepsilon e^{-st} \sum_{j=1}^{\infty} \frac{\lambda_j}{e^{-t} + \lambda_j} |b_j| \, |c_j| \,,$$

with the coefficients $b_j := (w, \psi_j)$ and $c_j := (\varphi, \psi_j)$. For the case $t > 0$, since $\frac{\lambda_j}{e^{-t} + \lambda_j} < 1$,

$$|\Psi(t + iy; w, \varphi)| \leq K_\varepsilon e^{-st} \|w\|_{L^2} \|\varphi\|_{L^2} \,. \tag{3.22}$$

When $t \leq 0$, a Young's inequality together with the assumption $\delta \in (s, 2-s)$ yield

$$\begin{aligned}
|\Psi(t+iy; w, \varphi)| &\leq K_\varepsilon e^{t(\delta-s)/2} \sum_{j=1}^{\infty} \frac{\lambda_j^{1-s/2-\delta/2} e^{-t(\delta+s)/2}}{e^{-t}+\lambda_j} \lambda_j^{\delta/2} |b_j| \lambda_j^{s/2} |c_j| \\
&\leq K_\varepsilon e^{t(\delta-s)/2} \sum_{j=1}^{\infty} \underbrace{\frac{\frac{2-s-\delta}{2}\lambda_j + \frac{\delta+s}{2} e^{-t}}{e^{-t}+\lambda_j}}_{\leq 1} \lambda_j^{\delta/2} |b_j| \lambda_j^{s/2} |c_j| \qquad (3.23) \\
&\leq K_\varepsilon e^{t(\delta-s)/2} \|w\|_{\dot{H}^\delta} \|\varphi\|_{\dot{H}^s}.
\end{aligned}$$

Estimates (3.22) and (3.23) yield (b) with

$$C = \begin{cases} K_\varepsilon (2\pi - 2\varepsilon) \|w\|_{L^2} \|\varphi\|_{L^2}, & t > 0, \\ K_\varepsilon (2\pi - 2\varepsilon) \|w\|_{\dot{H}^\delta} \|\varphi\|_{\dot{H}^s}, & t \leq 0, \end{cases}$$

and (c):

$$N(\Psi, \mathcal{D}_b) \leq 2K_\varepsilon (\delta - s)^{-1} \|w\|_{\dot{H}^\delta} \|\varphi\|_{\dot{H}^s} + K_\varepsilon s^{-1} \|w\|_{L^2} \|\varphi\|_{L^2} < \infty.$$

Proposition 2.3 now applies and guarantees that

$$\left| \int_{-\infty}^{\infty} \Psi(t; w, \varphi)\, dt - k \sum_{j=-\infty}^{\infty} \Psi(t_j; w, \varphi) \right| \leq \frac{N(\Psi, \mathcal{D}_b)}{\sinh(\pi(\pi-\varepsilon)/k)} e^{-\pi(\pi-\varepsilon)/k}. \qquad (3.24)$$

Furthermore, (3.22) and (3.23) also yield the following two tail estimates

$$k \sum_{j > N_+} \Psi(t_j; w, \varphi) \leq K_\varepsilon s^{-1} e^{-N_+ ks} \|w\|_{L^2} \|\varphi\|_{L^2}, \qquad (3.25a)$$

$$k \sum_{j < -N_-} \Psi(t_j; w, \varphi) \leq 2K_\varepsilon (\delta - s)^{-1} e^{-N_- k(\delta-s)/2} \|w\|_{\dot{H}^\delta} \|\varphi\|_{\dot{H}^s}. \qquad (3.25b)$$

The desired result follows from (3.24), (3.25a), and (3.25b) together with a triangle inequality. □

Remark 3.6 (Balancing the error)**.** Observe that asymptotically $\sinh(\pi^2/k - \pi\varepsilon/k) \sim e^{-\pi^2/k + \pi\varepsilon/k}$ as $k \to 0$. Therefore the first term in the error estimate (3.21) behaves like $e^{-2\pi^2/k + 2\pi\varepsilon/k}$. Thus assuming $\varepsilon \ll 1$ and setting $2\pi^2/k \approx N_+ ks \approx N_- k(\delta - s)/2$ balances the two error terms in

(3.21). More precisely, given a sinc quadrature step size $k > 0$, we set

$$N_+ = \left\lceil \frac{2\pi^2}{sk^2} \right\rceil, \quad \text{and } N_- = \left\lceil \frac{4\pi^2}{(\delta - s)k^2} \right\rceil$$

to simplify the sinc quadrature consistency error bound (3.21) as

$$\left|(a - a^k)(w, \varphi)\right| \leq K_\varepsilon \Big(s^{-1} \left\|w\right\|_{L^2} \left\|\varphi\right\|_{L^2} + 2(\delta - s)^{-1} \left\|w\right\|_{\dot{H}^\delta} \left\|\varphi\right\|_{\dot{H}^s} \Big) e^{-2\pi^2/k + 2\pi\varepsilon/k}. \tag{3.26}$$

In short, we write (3.26) as

$$\left|(a - a^k)(w, \varphi)\right| \leq C(s, \varepsilon, \delta) e^{-2\pi^2/k + 2\pi\varepsilon/k} \left\|w\right\|_{\dot{H}^\delta} \left\|\varphi\right\|_{\dot{H}^s}, \tag{3.27}$$

where the coefficient $C(s, \varepsilon, \delta) = K_\varepsilon(1/s + 2/(\delta - s))$ is independent of k.

3.3 Space discretization

The generic settings for the Galerkin finite element approximation is the same as in Section 2.5. The domain Ω is subdivided into a sequence of globally shape-regular, quasi-uniform, conforming subdivisions made of simplexes $\{\mathcal{T}_h : h > 0\}$. For any fixed $h > 0$, we denote V_h the space of piecewise continuous linear polynomials subordinate to $\mathcal{T}_h$ and vanish on $\partial\Omega$. With these settings we define the fully discrete bilinear form approximating (3.20), which reads

$$a_h^k(w, \varphi) := (\boldsymbol{u} \cdot \nabla w, \varphi) + \frac{c_s k}{2} \sum_{j=-N_-}^{N_+} e^{(1-s)t_j} \left((e^{t_j} + T_h)^{-1} \pi_h w, \varphi \right) \tag{3.28}$$

with w, $\varphi \in V$. We further define the finite element approximation to the weak problem (3.6) by: find $w_h \in V_h$ such that

$$a_h^k(w, \varphi_h) = (f, \varphi_h), \quad \text{for all } \varphi_h \in V_h. \tag{3.29}$$

The following lemma provides the consistency error between $a^k(\cdot, \cdot)$ and $a_h^k(\cdot, \cdot)$ on $V \times V_h$.

Lemma 3.7 (Finite element consistency error)**.** Let $s \in [1/2, 1)$. Let k be the sinc quadrature

stepping, and let N_+ and N_- be as in Remark 3.6. Let $a^k(w,\varphi)$ and $a_h^k(w,\varphi_h)$ be as in (3.20) and (3.28) respectively. Assume $\beta \in (s, 1+\alpha]$ and $\eta \in [s, \min(2s, 1+\alpha)]$. Then,

$$\left|(a^k - a_h^k)(w,\varphi_h)\right| \le C(h) h^{\min(2\alpha, \beta+\eta-2s)} \|w\|_{\dot{H}^\beta} \|\varphi_h\|_{\dot{H}_h^\eta}, \tag{3.30}$$

for all $w \in \dot{H}^\beta$ and $\varphi_h \in V_h$. Here for a constant C independent of h and k,

$$C(h) = \begin{cases} C, & \text{if } \beta - s > 2\alpha, \\ C\ln(1/h), & \text{if } \beta - s \le 2\alpha. \end{cases} \tag{3.31}$$

Proof. Throughout the proof the constant C is independent of h, k and t_j. Observe that from (3.20), (3.28) and the definitions of T and T_h, we have for any $w \in V$ and $\varphi_h \in V_h$,

$$(a^k - a_h^k)(w,\varphi_h) = \frac{c_s k}{2} \sum_{j=-N_-}^{N_+} e^{(1-s)t_j} \left([(e^{t_j}+T)^{-1} - (e^{t_j}+T_h)^{-1}\pi_h] w, \varphi_h \right).$$

Recalling the definition of the $L^2(\Omega)$ projection π_h in (2.19), we introduce

$$G(t) := (e^t+T)^{-1} - (e^t+T_h)^{-1}\pi_h = \pi_h(e^t+T)^{-1} - (e^t+T_h)^{-1}\pi_h.$$

Therefore, we write

$$\begin{aligned} (a^k - a_h^k)(w,\varphi_h) &= \frac{c_s k}{2} \sum_{j=-N_-}^{N_+} e^{(1-s)t_j} \left(G(t_j)w, \varphi_h\right) \\ &\le \frac{c_s k}{2} \sum_{t_j \le 0} e^{(1-s)t_j} \left|(G(t_j)w, \varphi_h)\right| + \frac{c_s k}{2} \sum_{t_j > 0} e^{(1-s)t_j} \left|(G(t_j)w, \varphi_h)\right| =: I + II. \end{aligned}$$

Because π_h and T_h commute, we can further write

$$\begin{aligned} G(t_j) &= (e^{t_j}+T_h)^{-1}(e^{t_j}+T_h)\pi_h(e^{t_j}+T)^{-1} - (e^{t_j}+T_h)^{-1}\pi_h(e^{t_j}+T)(e^{t_j}+T)^{-1} \\ &= (e^{t_j}+T_h)^{-1}\pi_h(T_h - T)(e^{t_j}+T)^{-1}. \end{aligned} \tag{3.32}$$

1 We start to estimate II by invoking the identity (3.32) to get

$$\begin{aligned}|(G(t_j)w,\varphi_h)| \leq & e^{-2t_j}\left\|(I+e^{-t_j}T_h)^{-1}\right\|_{L^2\to L^2}\|\pi_h(T-T_h)\|_{L^2\to L^2}\\ & \left\|(I+e^{-t_j}T)^{-1}\right\|_{L^2\to L^2}\|w\|_{L^2}\|\varphi_h\|_{L^2}.\end{aligned}$$

Using $\varphi=\sum\varphi_i\psi_i$ with $\varphi_i=\int_\Omega\varphi\psi_i$, and $\varphi_h=\sum\varphi_{h,i}\psi_{h,i}$ with $\varphi_{h,i}=\int_\Omega\varphi_h\psi_{h,i}$ we deduce that

$$\left\|(I+e^{-t_j}T_h)^{-1}\right\|_{L^2\to L^2}\leq 1,\quad \text{and } \left\|(I+e^{-t_j}T)^{-1}\right\|_{L^2\to L^2}\leq 1.$$

Furthermore, the L^2-stability of π_h and error estimate (2.18) with $r_1=r_2=\alpha$ yield

$$\|\pi_h(T-T_h)\|_{L^2\to L^2}\precsim\|T-T_h\|_{L^2\to L^2}\precsim h^{2\alpha}. \tag{3.33}$$

Gathering the above inequalities, we arrive at

$$\begin{aligned}II &\precsim \frac{c_s k}{2}h^{2\alpha}\sum_{t_j>0}e^{(1-s)t_j-2t_j}\|w\|_{L^2}\|\varphi_h\|_{L^2}\\ &\precsim \frac{c_s}{2}h^{2\alpha}\left(k\sum_{j>0}e^{-(1+s)jk}\right)\|w\|_{L^2}\|\varphi_h\|_{L^2}\precsim h^{2\alpha}\|w\|_{L^2}\|\varphi_h\|_{L^2}.\end{aligned}$$

2 To estimate I, we distinguish four cases depending on the regulative values of s, η, α and β. We first introduce $\epsilon:=\min(1-s,1/\ln(1/h))$, so that

$$\epsilon^{-1}\leq c_1\ln(1/h),\quad \text{and } h^{-\epsilon}\leq c_2 \tag{3.34}$$

for constants c_1 and c_2 independent of h.

Case 1: $2s-\eta\leq 1-\alpha$ and $\beta+\eta-2s>2\alpha$. We use the identity (3.32) and write

$$\begin{aligned}|(G(t_j)w,\varphi_h)| \leq & \left\|(e^{t_j}+T_h)^{-1}\right\|_{\dot H_h^{1-\alpha}\to\dot H_h^{-\eta}}\|\pi_h(T-T_h)\|_{\dot H^{\alpha-1}\to\dot H_h^{1-\alpha}}\\ & \left\|(e^{t_j}+T)^{-1}\right\|_{\dot H^{\beta}\to\dot H^{\alpha-1}}\|w\|_{\dot H^\beta}\|\varphi_h\|_{\dot H_h^\eta}.\end{aligned}$$

The three operator norms are estimated separately. We first note that

$$\left\|(e^{t_j}+T_h)^{-1}\right\|_{\dot{H}_h^{1-\alpha}\to\dot{H}_h^{-\eta}} \le e^{t_j(\eta-\alpha-1)/2},$$

using (2.17) with $\mu = e^{t_j}$, $a = (1-\alpha)/2$, $b = \eta/2$ ($\eta \le 1+\alpha$). Similarly, we have

$$\left\|(e^{t_j}+T)^{-1}\right\|_{\dot{H}^{\beta}\to\dot{H}^{\alpha-1}} \le e^{t_j(\beta-\alpha-1)/2},$$

using (2.13) with $\mu = e^{t_j}$, $a = \beta/2$, $b = (1-\alpha)/2$ ($\beta \le 1+\alpha$). In addition, the stability of π_h from $\dot{H}^{1-\alpha}$ to $\dot{H}_h^{1-\alpha}$, the norm equivalence (2.16) together with the error estimate (2.18) imply

$$\|\pi_h(T-T_h)\|_{\dot{H}^{\alpha-1}\to\dot{H}_h^{1-\alpha}} \le C\,\|T-T_h\|_{\dot{H}^{\alpha-1}\to\dot{H}^{1-\alpha}} \le Ch^{2\alpha}.$$

Gathering the above three estimates, we obtain

$$\begin{aligned} I &\le \frac{c_s C}{2} h^{2\alpha} k \sum_{t_j\le 0} e^{t_j(2-2s+\eta-\alpha-1+\beta-\alpha-1)/2}\,\|w\|_{\dot{H}^{\beta}}\,\|\varphi_h\|_{\dot{H}_h^{\eta}} \\ &= \frac{c_s C}{2} h^{2\alpha}\left(k\sum_{j\le 0} e^{kj(\beta-s-2\alpha+\eta-s)/2}\right)\|w\|_{\dot{H}^{\beta}}\,\|\varphi_h\|_{\dot{H}_h^{\eta}} \precsim h^{2\alpha}\,\|w\|_{\dot{H}^{\beta}}\,\|\varphi_h\|_{\dot{H}_h^{\eta}}, \end{aligned}$$

because $\beta+\eta-2s-2\alpha > 0$.

Case 2: $2s-\eta \le 1-\alpha$ and $0 < \beta+\eta-2s \le 2\alpha$. We set

$$\tilde{\alpha} := \frac{1}{2}(\beta+\eta-2s-\epsilon), \quad \tilde{\beta} = \min(\beta, 1+\tilde{\alpha}), \quad \text{and } \tilde{\eta} = \min(\eta, 1+\tilde{\alpha}),$$

so that $\tilde{\beta}+\tilde{\eta}-2s > 2\tilde{\alpha}$ and $2s-\tilde{\eta} \le 1-\tilde{\alpha}$. Repeating the argument in Case 1 by replacing α, β, η with $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\eta}$ respectively yields

$$I \le C\epsilon^{-1}h^{2\tilde{\alpha}}\,\|w\|_{\dot{H}^{\tilde{\beta}}}\,\|\varphi_h\|_{\dot{H}_h^{\tilde{\eta}}} \le C\epsilon^{-1}h^{\beta+\eta-2s-\epsilon}\,\|w\|_{\dot{H}^{\beta}}\,\|\varphi_h\|_{\dot{H}_h^{\eta}}.$$

Case 3: $1-\alpha < 2s-\eta \le s$ and $1 \le \beta \le 1+\alpha$. Using (2.17) with $\mu = e^{t_j}$, $a = (2s-\eta+\epsilon)/2$, $b = \eta/2$; (2.13) with $\mu = e^{t_j}$, $a = \beta/2$, $b = (2-\beta)/2$; and (2.18) with $r_1 = 1-2s+\eta-\epsilon$ and $r_2 = \beta - 1$, we obtain

$$\begin{aligned}
|(G(t_j)w, \varphi_h)| &\le \left\|(e^{t_j}+T_h)^{-1}\right\|_{\dot{H}_h^{2s-\eta+\epsilon}\to\dot{H}_h^{-\eta}} \left\|\pi_h(T-T_h)\right\|_{\dot{H}^{\beta-2}\to\dot{H}_h^{2s-\eta+\epsilon}} \\
&\quad \left\|(e^{t_j}+T)^{-1}\right\|_{\dot{H}^{\beta}\to\dot{H}^{\beta-2}} \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}} \\
&\le C e^{t_j(2s+\epsilon-2)/2} h^{\beta+\eta-2s-\epsilon} \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}},
\end{aligned}$$

so that

$$\begin{aligned}
I &\le \frac{c_s C}{2} h^{\beta+\eta-2s-\epsilon} k \sum_{t_j\le 0} e^{t_j(2-2s+2s-2+\epsilon)/2} \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}} \\
&= \frac{c_s C}{2} h^{\beta+\eta-2s-\epsilon} \left(k \sum_{j\le 0} e^{kj\epsilon/2}\right) \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}} \precsim \epsilon^{-1} h^{\beta+\eta-2s-\epsilon} \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}}.
\end{aligned}$$

Case 4: $1-\alpha < 2s-\eta \le s$ and $s < \beta < 1$. Using (2.17) with $\mu = e^{t_j}$, $a = (1-\beta+2s-\eta+\epsilon)/2$, $b = \eta/2$; (2.13) with $\mu = e^{t_j}$, $a = \beta/2$, $b = 1/2$; and (2.18) with $r_1 = \beta - 2s + \eta - \epsilon$ and $r_2 = 0$, we have

$$\begin{aligned}
|(G(t_j)w, \varphi_h)| &\le \left\|(e^{t_j}+T_h)^{-1}\right\|_{\dot{H}_h^{1-\beta-\eta+2s+\epsilon}\to\dot{H}_h^{-\eta}} \left\|\pi_h(T-T_h)\right\|_{\dot{H}^{-1}\to\dot{H}_h^{1-\beta-\eta+2s+\epsilon}} \\
&\quad \left\|(e^{t_j}+T)^{-1}\right\|_{\dot{H}^{\beta}\to\dot{H}^{-1}} \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}} \\
&\le C\epsilon^{-1} e^{t_j(-\beta+2s+\epsilon-1)/2} h^{\beta+\eta-2s-\epsilon} e^{t_j(\beta-1)} \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}},
\end{aligned}$$

and therefore

$$\begin{aligned}
I &\le \frac{c_s C}{2} h^{\beta+\eta-2s-\epsilon} k \sum_{t_j\le 0} e^{t_j(2-2s-\beta+2s+\epsilon-1+\beta-1)/2} \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}} \\
&= \frac{c_s C}{2} h^{\beta+\eta-2s-\epsilon} \left(k \sum_{j\le 0} e^{kj\epsilon/2}\right) \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}} \precsim \epsilon^{-1} h^{\beta+\eta-2s-\epsilon} \|w\|_{\dot{H}^{\beta}} \|\varphi_h\|_{\dot{H}_h^{\eta}}.
\end{aligned}$$

3 Invoking (3.34) in Cases 2, 3 and 4 shows that

$$C\epsilon^{-1}h^{\beta+\eta-2s-\epsilon} \leq C(h)h^{\beta+\eta-2s}.$$

Combining the estimates I and II yields the result. □

The consistency error between $a(\cdot,\cdot)$ and $a_h^k(\cdot,\cdot)$ on $V \times V_h$ follows from Lemmas 3.5 and 3.7. This is the object of the theorem below.

Theorem 3.8. Let $s \in [1/2, 1)$. Let $\alpha \in (0, 1]$ such that Assumption 2.2 holds. Let $\varepsilon \ll 1$ be the constant in Lemma 3.5. Let N_+ and N_- defined in Remark 3.6 with sinc quadrature step size k given. Let $\{\mathcal{T}_h\}$ be a sequence of shape-regular subdivisions of Ω. Suppose $a(w,\varphi)$ and $a_h^k(w,\varphi_h)$ are defined as in (3.6) and (3.28) respectively and assume $w \in \mathbb{H}^\beta$ for $\beta \in (s, 1+\alpha]$. Then

$$\left|(a - a_h^k)(w,\varphi_h)\right| \leq C\left(e^{-2\pi^2/k+2\pi\varepsilon/k} + C(h)h^{\min(2\alpha,\beta-s)}\right)\|w\|_{\mathbb{H}^\beta}\|\varphi_h\|_{\mathbb{H}^s}, \tag{3.35}$$

where $C(h)$ is the constant appearing in Lemma 3.7.

Proof. For $\delta \in (s, \min(2-s, \beta)]$, (3.27) guarantees that

$$\left|(a - a^k)(w,\varphi_h)\right| \leq C(s,\delta)e^{-2\pi^2/k+2\pi\varepsilon/k}\|w\|_{\dot{H}^\delta}\|\varphi_h\|_{\dot{H}^s} \leq C(s,\delta)e^{-2\pi^2/k+2\pi\varepsilon/k}\|w\|_{\dot{H}^\beta}\|\varphi_h\|_{\dot{H}^s}.$$

Choosing $\eta = s$ in (3.30) together with the norm equivalence (2.16) yield

$$\left|(a_h - a_h^k)(w,\varphi_h)\right| \leq C(h)h^{\min(2\alpha,\beta-s)}\|w\|_{\dot{H}^\beta}\|\varphi_h\|_{\dot{H}_h^s} \lesssim C(h)h^{\min(2\alpha,\beta-s)}\|w\|_{\dot{H}^\beta}\|\varphi_h\|_{\dot{H}^s}.$$

It suffices to invoke a triangle inequality and the norm equivalence (2.7) to complete the proof. □

Remark 3.9 (Balance the errors)**.** The two terms in error estimate (3.35) are balanced when $e^{-2\pi^2/k} \approx h^{\min(2\alpha,\beta-s)}$. That is, given the mesh size h, setting $k = \frac{\pi^2}{\min(2\alpha,\beta-s)\ln(1/h)}$ yields

$$\left|(a - a_h^k)(w,\varphi_h)\right| \leq C(h)h^{\min(2\alpha,\beta-s)}\|w\|_{\mathbb{H}^\beta}\|\varphi_h\|_{\mathbb{H}^s}. \tag{3.36}$$

3.4 Energy and L^2 error estimates

The analysis of the fully discrete scheme relies on L^2 error estimates for the stationary problem. They are obtained by the duality argument using error estimates in the energy norm. The latter is based on a Strang's lemma to account for the consistency error generated by the approximation of the operator. We start with the V_h-ellipticity and continuity.

Lemma 3.10 (V_h-ellipticity and continuity)**.** Let $\delta \in (s, 2-s]$, k be the sinc quadrature stepping and the parameters N_+, N_-, and $C(s,\varepsilon,\delta)$ be given as in Remark 3.6. Let $\varepsilon \ll 1$ be the constant in Lemma 3.5. Let $\{\mathcal{T}_h\}$ be a sequence of shape-regular subdivisions of Ω. Suppose $k = k(h)$ is chosen sufficiently small such that

$$C_I C(s,\varepsilon,\delta) e^{-2\pi^2/k + 2\pi\varepsilon/k} h^{s-\delta} := D_s < 1, \tag{3.37}$$

with C_I the constant in inverse inequality (2.22), then there exist constants C_1 and C_2 independent of h and k such that for all $w \in V$, $\varphi_h \in V_h$

$$a_h^k(\varphi_h, \varphi_h) \geq C_1 \left\|\varphi_h\right\|_{\dot{H}^s}^2, \qquad \text{and} \qquad a_h^k(w, \varphi_h) \leq C_2 \left\|w\right\|_{\dot{H}^s} \left\|\varphi_h\right\|_{\dot{H}^s}.$$

Proof. Without losing of generality we assume $h < 1$. The V_h-ellipticity was proved in [27] Theorem 7.2. However, we emphasis that the constant

$$C_1 = 1 - C_I C(s,\varepsilon,\delta) e^{-2\pi^2/k + 2\pi\varepsilon/k} h^{s-\delta} = 1 - D_s > 0.$$

The continuity of $a_h^k(\cdot,\cdot)$ relies on the continuity of

$$a_h(w, \varphi_h) := \langle \boldsymbol{u} \cdot \nabla w, \varphi \rangle + (L_h^{s/2} \pi_h w, L_h^{s/2} \varphi_h),$$

defined for $w \in V$ and $\varphi_h \in V_h$. In fact, from the norm equivalence inequality (2.16) and stability

of π_h, we obtain that

$$|a_h(w, \varphi_h)| \leq C \|w\|_{\dot{H}^s} \|\varphi_h\|_{\dot{H}^s} ,$$

where C is a constant independent of h. In addition, the argumentation in the proof of Lemma 3.5 extends readily when L is replaced by L_h. Therefore we have

$$\left|(a_h - a_h^k)(w, \varphi_h)\right| \leq C(s, \varepsilon, \delta) e^{-2\pi^2/k + 2\pi\varepsilon/k} \|w\|_{\dot{H}^\delta} \|\varphi_h\|_{\dot{H}_h^s} ,$$

where $C(s, \delta)$ is as in Remark 3.6. A triangle inequality and the inverse estimate (2.23) are then employed to write

$$\begin{aligned} a_h^k(w, \varphi_h) &\leq \left|(a_h - a_h^k)(w, \varphi_h)\right| + |a_h(w, \varphi_h)| \\ &\leq C_I C(s, \varepsilon, \delta) e^{-2\pi^2/k + 2\pi\varepsilon/k} h^{s-\delta} \|w\|_{\dot{H}^s} \|\varphi_h\|_{\dot{H}^s} + C \|w\|_{\dot{H}^s} \|\varphi_h\|_{\dot{H}^s} \\ &\leq (C + D_s) \|w\|_{\dot{H}^s} \|\varphi_h\|_{\dot{H}^s} . \end{aligned}$$

This completes the proof. □

Lemma 3.11 (Energy error)**.** Let $s \in [1/2, 1)$ and α be the elliptic pickup regularity index in Assumption 2.2. Let $\varepsilon \ll 1$ be the constant in Lemma 3.5. Let k be the sinc quadrature step, and let N_+ and N_- be as in Remark 3.6. For a given mesh size h, suppose k is chosen sufficiently small such that (3.37) holds. Let $w \in V$ and $w_h \in V_h$ be the solutions to problems (3.6) and (3.29) respectively and assume $w \in \mathbb{H}^\beta$ for $\beta \in (s, 1 + \alpha]$. Then

$$\|w - w_h\|_{\mathbb{H}^s} \leq C \left(e^{-2\pi^2/k + 2\pi\varepsilon/k} + C(h) h^{\min(2\alpha, \beta - s)} \right) \|w\|_{\mathbb{H}^\beta} , \tag{3.38}$$

where $C(h)$ is defined by (3.31).

Proof. Throughout this proof C is a generic constant independent of h and k. Thanks to the V_h ellipticity and the continuity guaranteed by Lemma 3.10, we have the following second Strang's

lemma

$$\begin{aligned}\|w-w_h\|_{\dot{H}^s} &\leq C\Big(\inf_{v_h\in V_h}\|w-v_h\|_{\dot{H}^s}+\sup_{\chi\in V_h}\frac{|(a-a_h^k)(w,\chi)|}{\|\chi\|_{\dot{H}^s}}\Big)\\ &\leq C\Big(\|w-\pi_h w\|_{\mathbb{H}^s}+\sup_{\chi\in V_h}\frac{|(a-a_h^k)(w,\chi)|}{\|\chi\|_{\mathbb{H}^s}}\Big),\end{aligned} \tag{3.39}$$

where in the last inequality we applied the norm equivalence (2.7). Recall that by assumption $\min(2\alpha,\beta-s)+s\leq\beta$. Therefore, the L^2 projection error estimate (2.20) yields

$$\|w-\pi_h w\|_{\mathbb{H}^s}\lesssim h^{\min(2\alpha,\beta-s)}\|w\|_{\mathbb{H}^{\min(2\alpha,\beta-s)+s}}\lesssim h^{\min(2\alpha,\beta-s)}\|w\|_{\mathbb{H}^\beta}.$$

The above estimate together with consistency error (3.35) and the norm equivalence (2.7) ends the proof of (3.38). □

A duality argument is put forward to derive the L^2 error estimates. In this aim, consider the following problem. Find $z\in V$ such that

$$a^*(z,\varphi):=-\langle\boldsymbol{u}\cdot\nabla z,\varphi\rangle+(L^{s/2}z,L^{s/2}\varphi)=(w-w_h,\varphi),\ \text{ for all }\varphi\in V. \tag{3.40}$$

By definition, we have the following important identity

$$a^*(\chi,w-w_h)=a(w-w_h,\chi)=(a_h^k-a)(w_h,\chi),\ \text{ for all }\chi\in V_h. \tag{3.41}$$

Remark 3.12. Upon replacing $\boldsymbol{u}$ by $-\boldsymbol{u}$ in (3.1), the dual problem (3.40) has the same regularity as (3.1). Namely, for $\boldsymbol{u}\in\mathbf{U}$, (3.40) is well-posed, and $a^*(\cdot,\cdot):V\times V\to\mathbb{R}$ is coercive and continuous (see Section 3.1). In addition, Proposition 3.1 translates to

$$\|z\|_{\dot{H}^{2s^*}}\leq C\|w-w_h\|, \tag{3.42}$$

where the constant C is independent of s and h, and s^* is defined by (3.8).

We are now in position to derive the L^2 error estimates.

Theorem 3.13 (L^2 error)**.** Let $s \in [1/2, 1)$, s^* be the constant defined by (3.8), and α be the elliptic pickup regularity index in Assumption 2.2. Let $\varepsilon \ll 1$ be the constant in Lemma 3.5. Let k be the sinc quadrature stepping, and let N_+ and N_- be as in Remark 3.6. For a given mesh size h, suppose k is chosen sufficiently small such that (3.37) holds. Let $w \in V$ and $w_h \in V_h$ be the solutions to problems (3.6) and (3.29) respectively and assume $w \in \mathbb{H}^\beta$ for $\beta \in (s, 1+\alpha]$. Then

$$\|w - w_h\| \leq C(h) \left(e^{-2\pi^2/k + 2\pi\varepsilon/k} + C(h) h^{\min(2\alpha, \beta + 2s^* - 2s, \beta + 1 + \alpha - 2s)} \right) \|w\|_{\mathbb{H}^\beta}, \tag{3.43}$$

where $C(h)$ is given by (3.31).

Proof. When $\beta - s > 2\alpha$, Lemma 3.11 automatically ensures that

$$\|w - w_h\| \leq \|w - w_h\|_{\dot{H}^s} \leq C(h) \left(e^{-2\pi^2/k + 2\pi\varepsilon/k} + h^{2\alpha} \right) \|w\|_{\mathbb{H}^\beta}.$$

Therefore, we now focus on the case $\beta - s \leq 2\alpha$, i.e., when $\min(2\alpha, \beta - s) = \beta - s$. For this we use the dual problem (3.40) and equality (3.41) to write

$$\begin{aligned} \|w - w_h\|^2 &= a^*(z, w - w_h) = a^*(z - \pi_h z, w - w_h) + (a_h^k - a)(w_h, \pi_h z) \\ &= a^*(z - \pi_h z, w - w_h) + (a_h^k - a)(w_h - \pi_h w, \pi_h z) + (a_h^k - a)(\pi_h w, \pi_h z) \\ &=: E_1 + E_2 + E_3. \end{aligned}$$

We set $\bar{\beta} = \min(2\alpha, \beta + 2s^* - 2s, \beta + 1 + \alpha - 2s) - (\beta - s) + s$, then

$$\bar{\beta} = \begin{cases} 2\alpha - \beta + 2s, & \text{if } \beta + \min(2s^*, 1 + \alpha) - 2s \geq 2\alpha, \\ \min(2s^*, 1 + \alpha), & \text{if } \beta + \min(2s^*, 1 + \alpha) - 2s < 2\alpha. \end{cases}$$

Therefore it directly follows from the definition that $s < \bar{\beta} \leq \min(2s^*, 1 + \alpha)$, $\bar{\beta} - s \leq 2\alpha$, and $\bar{\beta} + \beta - 2s = \min(2\alpha, \beta + 2s^* - 2s, \beta + 1 + \alpha - 2s)$. We first apply the continuity of $a^*(\cdot, \cdot)$,

the approximation property of π_h (2.20) and the regularity of the dual problem (3.42) together with Lemma 3.11 to bound E_1,

$$\begin{aligned}
E_1 &\leq C \left\|(I-\pi_h)z\right\|_{\dot{H}^s} \left\|w-w_h\right\|_{\dot{H}^s} \\
&\leq Ch^{\bar{\beta}-s}\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\beta-s}\right)\left\|z\right\|_{\dot{H}^{\bar{\beta}}}\left\|w\right\|_{\mathbb{H}^\beta} \\
&\leq C\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\min(2\alpha,\beta+2s^*-2s,\beta+1+\alpha-2s)}\right)\left\|w\right\|_{\mathbb{H}^\beta}\left\|w-w_h\right\|.
\end{aligned}$$

To bound E_2, we first infer from the triangle inequality, the energy error estimate (3.38) and the L^2-projection estimate (2.20) that

$$\begin{aligned}
\left\|w_h-\pi_h w\right\|_{\dot{H}^s} &\leq \left\|w_h-w\right\|_{\dot{H}^s}+\left\|w-\pi_h w\right\|_{\dot{H}^s} \\
&\leq C\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\beta-s}\right)\left\|w\right\|_{\dot{H}^\beta}+Ch^{\beta-s}\left\|w\right\|_{\dot{H}^\beta} \\
&\leq C\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\beta-s}\right)\left\|w\right\|_{\dot{H}^\beta}.
\end{aligned}$$

Hence it follows additionally from the consistency error estimate (3.35), the stability of π_h (2.21) and the regularity of the dual problem (3.42) that

$$\begin{aligned}
E_2 &\leq C\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\bar{\beta}-s}\right)\left\|\pi_h z\right\|_{\dot{H}^{\bar{\beta}}}\left\|w_h-\pi_h w\right\|_{\dot{H}^s} \\
&\leq C\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\bar{\beta}-s}\right)\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\beta-s}\right)\left\|\pi_h z\right\|_{\dot{H}^{\bar{\beta}}}\left\|w\right\|_{\dot{H}^\beta} \\
&\leq C\ln(1/h)\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\min(2\alpha,\beta+2s^*-2s,\beta+1+\alpha-2s)}\right)\left\|w\right\|_{\mathbb{H}^\beta}\left\|w-w_h\right\|,
\end{aligned}$$

because $\bar{\beta}-s<2\alpha$ and $\bar{\beta}\leq\min(2s^*,1+\alpha)$.

Invoking (3.35) again by taking $\eta=\bar{\beta}$ we obtain the bound for E_3,

$$\begin{aligned}
E_3 &\leq C\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\beta+\bar{\beta}-2s}\right)\left\|\pi_h w\right\|_{\dot{H}^\beta}\left\|\pi_h z\right\|_{\dot{H}_h^{\bar{\beta}}} \\
&\precsim C\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{\min(2\alpha,\beta+2s^*-2s,\beta+1+\alpha-2s)}\right)\left\|w\right\|_{\mathbb{H}^\beta}\left\|w-w_h\right\|.
\end{aligned}$$

The proof is completed after collecting estimates E_1, E_2 and E_3 and canceling the common term $\|w - w_h\|$ throughout. $\square$

3.5 Numerical illustration

As the sinc quadrature error has already been thoroughly discussed in [23, 25, 29, 27], in all the following cases we choose k small enough that the influence of sinc quadrature error is neglectable.

3.5.1 Regularity for $s = 1/2$

To check the regularity degeneration for the critical case $s = 1/2$, we set $\Omega = [0, 1]^2$, $\boldsymbol{u} = (U_0, 0)$ with U_0 and mesh size h the varying variables. It is well-known that when the advection term dominates (i.e., U_0 large), the Galerkin approximation will result in spurious oscillations, thus a stabilization regime is required. We apply the second order maximum preserving viscosity introduced in [36]. We take the right hand side to be

$$f(x) = \lambda^s \sin(\pi n x_1) \sin(\pi m x_2). \tag{3.44}$$

Here $m,\ n \in \mathbb{N}$, and $\lambda = \pi^2(m^2+n^2)$. We choose $m = n = 1$ and report in Figure 3.1 the solution cut over by the plane $\{y = 0.5\}$.

3.5.2 Finite element error for stationary problem

Set the domain $\Omega = [0, 1]^2$, velocity $\boldsymbol{u} = (U_1, U_2)$. When the data f has enough regularity such that the solution $w \in \dot{H}^\beta$ with $\beta \geq \alpha = 1$, we expect a L^2 error convergence proportional to h^2, which is exactly the case when the exact solution is chosen to be (3.44). We test with $m = 1,\ n = 2$, $s = 0.5,\ 0.75$, $\boldsymbol{u} = (1, 1)$, and with different mesh sizes to report in Table 3.1 the observed convergence rate by

$$ORC_i = \ln(e_i/e_{i+1})/\ln(e_i/e_{i+1}).$$

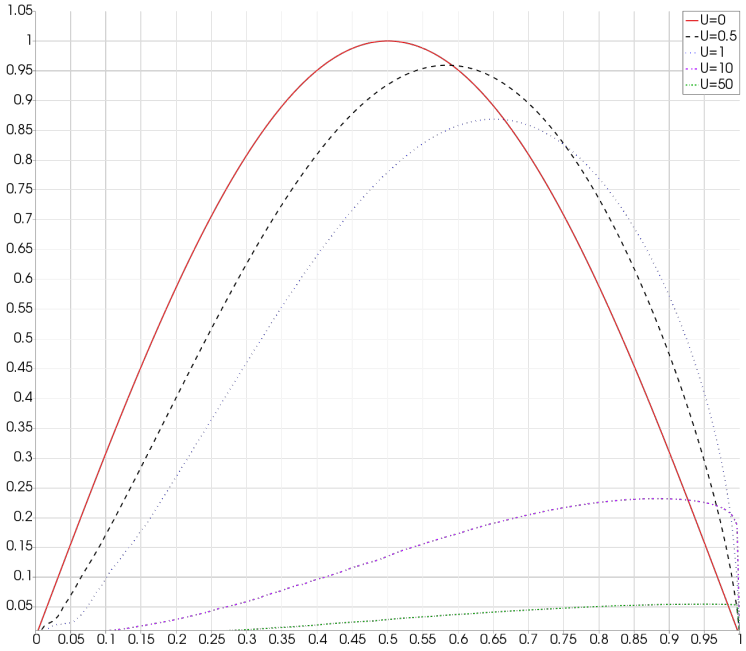


Figure 3.1. The cut of solution over the plane $\{y = 0.5\}$ with different values of U_0. Mesh refined 8 times.

	$s = 0.5$		$s = 0.75$	
h	L^2 error	ORC	L^2 error	ORC
1/4	0.0712736		0.0835161	
1/8	1.531E-02	2.22	1.997E-02	2.06
1/16	3.694E-03	2.05	4.941E-03	2.02
1/32	9.153E-04	2.01	1.232E-03	2.00
1/64	2.283E-04	2.00	3.078E-04	2.00
1/128	5.706E-05	2.00	7.691E-05	2.00

Table 3.1. L^2 error and observed rate of convergence (ORC) with the right hand side taking smooth function. The theoretical convergence rate is 2.

We also consider the checker-board benchmark, i.e., set

$$g(x_1, x_2) = \begin{cases} 1, & \text{if } (x_1 - 0.5)(x_2 - 0.5) > 0, \\ -1, & \text{otherwise.} \end{cases} \tag{3.45}$$

A complete set of eigen-pairs of $L = -\Delta$ coupled with homogeneous boundary conditions on $\Omega = (0,1)^2$ is $\{\lambda_{mn},\, \psi_{mn}\}_{m,n=1}^{\infty}$, with $\lambda_{mn} = \pi^2(m^2 + n^2)$ and $\psi_{mn} = \sin(\pi m x_1)\sin(\pi n x_2)$. Therefore, the checker-board function has expansion $g = \sum_{m,n=1}^{\infty} c_{mn}\psi_{mn}$ for $c_{mn} = (g, \psi_{mn})$. We consider $\boldsymbol{u} = (U_1, U_2)$ and set the right hand side

$$f = g + \sum_{m,n=1}^{\infty} c_{mn}\lambda_{mn}^{-s}\boldsymbol{u}\cdot\nabla\psi_{mn}.$$

Notice that with respect to the right hand side data f, the exact solution is

$$w = \sum_{m,n=1}^{\infty} c_{mn}\lambda_{mn}^{-s}\psi_{mn} \in \dot{H}^{\frac{1}{2}+2s-\epsilon}$$

for any $\epsilon > 0$. Consequently, $\sum_{m,n=1}^{\infty} c_{mn}\lambda_{mn}^{-s}\boldsymbol{u}\cdot\nabla\psi_{mn} \in \dot{H}^{\frac{1}{2}-\epsilon}(\Omega)$, ensuring $f \in \dot{H}^{\frac{1}{2}-\epsilon}(\Omega)$. Theorem 3.13 predicts an error decay of order

$$\max(1, 1 + \ln(1/h))h^{\min(2, \frac{1}{2}+4s^*-2s-\epsilon)}.$$

In practice we truncate the series in f and w by 800 modes in each direction ($64,000$ modes in total). $U_1 = U_2$ is set to be 0, 0.1, 1, 10 respectively. The SINC quadrature points are taken large enough not to pollute the space discretization error. The average of the observed convergence rates are reported in Table 3.2 and are compared with theoretical rates predicted by Theorem 3.13.

We present in Figure 3.2 the cut of the plot over plane $\{y = 1/4\}$ for $s = 0.5,\, 0.7,\, 0.9$ with $\boldsymbol{u} = (1, 1)$.

s	0.6	0.7	0.8	0.9
THM	1.7	1.9	2.0	2.0
$U = 0$	1.66	1.81	1.91	1.97
$U = 0.1$	1.68	1.83	1.93	1.98
$U = 1$	1.66	1.81	1.91	1.97
$U = 10$	1.37	1.72	1.91	1.97

Table 3.2. L^2 error and observed rate of convergence (ORC) with the right hand side taking checker board function.

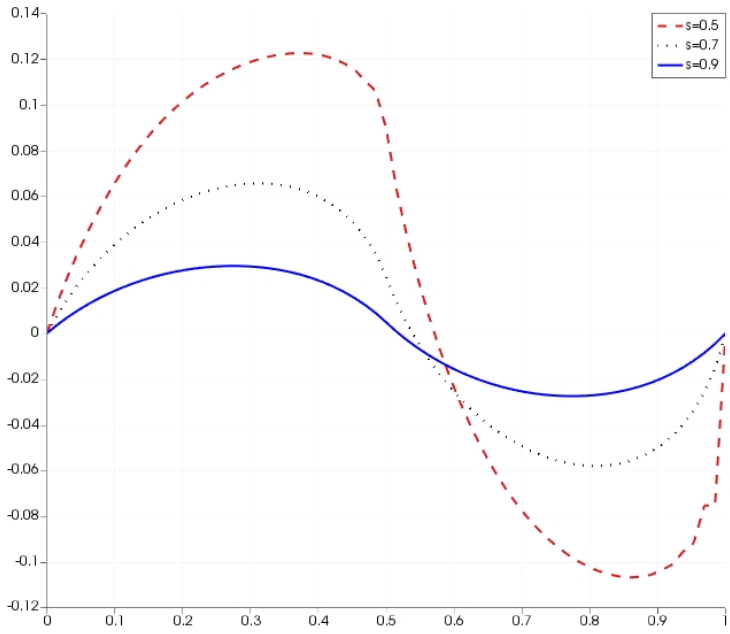


Figure 3.2. Cut-over view for solution with checker-board data.

4. APPROXIMATION OF THE TIME-DEPENDENT PROBLEM

The goal of this chapter is to develop a numerical approximation to the time-dependent problem (1.3). Following the definition of weak formulation for the stationary problem as in (3.6), we establish a weak formulation for the time-dependent problem (1.3) by: given $f \in L^2(0, \mathtt{T}; L^2(\Omega))$, $v \in L^2(\Omega)$ (at least, more regularity will be assumed in the error analysis if necessary) and $s \in [1/2, 1)$, find $q \in L^2(0, \mathtt{T}; \dot{H}^s(\Omega))$ with $\partial_t q \in L^2(0, \mathtt{T}; \dot{H}^{-s}(\Omega))$ such that

$$\langle \frac{\partial}{\partial t} q, \varphi \rangle + a(q, \varphi) = (f, \varphi), \qquad q(t = 0) = v, \quad \forall \varphi \in \dot{H}^s(\Omega), \tag{4.1}$$

for a.e. $t \in [0, \mathtt{T}]$, here the bilinear form $a(q, \varphi)$ is defined in (3.6).

In our discussions we consider two different settings to the initial and right hand side data: the homogeneous problem with $f = 0$ and $0 \neq v \in \mathbb{H}^{2\gamma}(\Omega)$ for $\gamma \in [0, (1 + \alpha)/2]$, and the non-homogeneous problem with $v = 0$ and $f \in L^\infty(0, \mathtt{T}; \mathbb{H}^{2\gamma}(\Omega))$.

We first derive the full discretization scheme in Section 4.1. It is achieved by first applying the space discretization scheme constructed in Chapter 3 followed by a forward Euler time stepping. The stability condition is provided therein. The total error estimates for homogeneous problem and non-homogeneous problem are presented in Sections 4.2 and **??** respectively. Numerical results are reported in Section 4.4 to support the error analysis for above discretization schemes.

4.1 Fully discretization scheme and stability requirement

We first define two operators associated to the weak problems (3.6) and (3.29). Define $\mathcal{S} : L^2(\Omega) \to V$ to be the solution operator to the problem

$$a(w, \varphi) = \langle \boldsymbol{u} \cdot \nabla w, \varphi \rangle + (L^{s/2} w, L^{s/2} \varphi) = (f, \varphi), \qquad \text{for all } \varphi \in V, \tag{4.2}$$

namely, $w = \mathcal{S} f$. Clearly $\mathcal{S}$ is one to one. We then define $\mathcal{A} = \mathcal{S}^{-1}$ with domain $D(\mathcal{A}) = \mathrm{Range}(\mathcal{S})$. By construction $\mathcal{A}$ is a regularly accretive operator (cf, [25]). In addition, Theorem 2.2

of [22] implies $\mathcal{A}$ is a sectorial operator. That is, there exists constant $\omega \in [0, \pi/2)$, such that the spectrum of $\mathcal{A}$ is subset of the sector $S_\omega := \{z \in \mathbb{C} : |\arg z| \leq \omega\}$.

The space Ω is discretized into a sequence of globally shape-regular, quasi-uniform and conforming subdivisions $\{\mathcal{T}_h : h > 0\}$ in the same way as in the stationary problem in Chapter 3. It allows us to define the discrete counterpart of $\mathcal{S}$ by $\mathcal{S}_h : V_h \to V_h$ to be the solution operator to the discrete problem

$$a_h^k(w_h, \varphi_h) = (\pi_h f, \varphi_h), \tag{4.3}$$

i.e., $w_h = \mathcal{S}_h \pi_h f$. We then define $\mathcal{A}_h := \mathcal{S}_h^{-1}$ accordingly.

Recalling the finite element approximation $a_h^k(q_h, \varphi_h)$ to the bilinear form $a(\cdot, \cdot)$ proposed in (3.29), we define the finite element approximation of (4.1) by: seek $q_h(t) \in H^1((0, \mathtt{T}); V_h)$, satisfying

$$\Big(\frac{\partial}{\partial t} q_h, \varphi_h\Big) + a_h^k(q_h, \varphi_h) = (f, \varphi_h), \quad q_h(t = 0) = v_h := \pi_h v, \qquad \text{for all } \varphi_h \in V_h \tag{4.4}$$

holds for a.e. $t \in [0, \mathtt{T}]$, with parameters N_+, N_- and $C(s, \delta)$ provided in Remark 3.6 assuming sinc quadrature stepping k given.

For the computational expenses consideration, we treat the bilinear form $a_h^k(q_h, \varphi_h)$ explicitly, and consider the time discretization to (4.4) with the forward Euler scheme. Given time stepping $\tau > 0$, we denote $t_n := n\tau$ for $n = 0, 1, 2, \cdots$. The resulting fully discretized problem reads: for the given $q_h^n \in V_h$ approximating $q(t_n)$, find $q_h^{n+1} \in V_h$ approximating $q(t_{n+1})$, such that

$$(q_h^{n+1}, \varphi_h) = (q_h^n, \varphi_h) - \tau a_h^k(q_h^n, \varphi_h) + \tau(f^n, \varphi_h), \qquad \text{for all } \varphi_h \in V_h. \tag{4.5}$$

Here $f^n := f(t_n)$, and the initial condition $q_h^0 = v_h := \pi_h v$ is given to allow the algorithm to start.

We end the discussion in this Section by showing that the forward Euler method is stable under a stability condition.

Lemma 4.1 (Stability)**.** Let $\delta \in (s, \min(3/2, 2 - s)]$, let k be the sinc quadrature step size and

constants N_+, N_- and $C(s,\delta)$ be given in Remark 3.6. Suppose k is chosen sufficiently small such that condition (3.37) holds with constant $D_s < 1/2$. Then the fully discretization scheme (4.5) is stable in $L^2(\omega)$ provided the stability condition $\tau \le Ch^{2s}$ for some constant C independent of h.

Proof. Throughout the proof the constant C is independent of h, k and t_j. Additionally, without loosing of generality, we assume mesh size $h < 1$. We denote by $\boldsymbol{u}^n := \boldsymbol{u}(t^n)$, and rewrite in (4.5) the bilinear term

$$a_h^k(q_h^n, \varphi_h) = (a_h^k - a)(q_h^n, \varphi_h) + (\boldsymbol{u}^n \cdot \nabla q_h^n, \varphi_h) + (L^{s/2} q_h^n, L^{s/2}\varphi_h).$$

We first put $\varphi_h = 2\tau q_h^n = -\tau(q_h^{n+1} - q_h^n) + \tau(q_h^{n+1} + q_h^n)$ into (4.5), which yields

$$\left\|q_h^{n+1}\right\|_{L^2}^2 - \|q_h^n\|_{L^2}^2 + 2\tau \left\|L^{s/2} q_h^n\right\|_{L^2}^2 = \left\|q_h^{n+1} - q_h^n\right\|_{L^2}^2 + 2\tau(f^n, q_h^n) - 2\tau(a_h^k - a)(q_h^n, q_h^n). \quad (4.6)$$

Next we set $\varphi_h = q_h^{n+1} - q_h^n$ in (4.5) to obtain

$$\begin{aligned} \frac{1}{\tau}\left\|q_h^{n+1} - q_h^n\right\|_{L^2}^2 &= (f^n, q_h^{n+1} - q_h^n) - (\boldsymbol{u}^n \cdot \nabla q_h^n, q_h^{n+1} - q_h^n) \\ &\quad - (L^{s/2} q_h^n, L^{s/2}(q_h^{n+1} - q_h^n)) + (a - a_h^k)(q_h^n, q_h^{n+1} - q_h^n) \qquad (4.7) \\ &=: B_1 + B_2 + B_3 + B_4. \end{aligned}$$

The first three terms B_1, B_2 and B_3 can be bounded by applying Cauchy-Schwarz inequality and the inverse inequality (2.23), that is,

$$\begin{aligned} &B_1 \le \|f^n\|_{L^2} \left\|q_h^{n+1} - q_h^n\right\|_{L^2}, \\ &B_2 \le C \|\boldsymbol{u}^n\|_{\mathbf{L}^\infty} \|\nabla q_h^n\|_{L^2} \left\|q_h^{n+1} - q_h^n\right\|_{L^2} \le Ch^{s-1} \|\boldsymbol{u}^n\|_{\mathbf{L}^\infty} \|q_h^n\|_{\dot{H}^s} \left\|q_h^{n+1} - q_h^n\right\|_{L^2}, \\ &B_3 \le \left\|L^{s/2} q_h^n\right\|_{L^2} \left\|L^{s/2}(q_h^{n+1} - q_h^n)\right\|_{L^2} = \|q_h^n\|_{\dot{H}^s} \left\|q_h^{n+1} - q_h^n\right\|_{\dot{H}^s} \le Ch^{-s} \|q_h^n\|_{\dot{H}^s} \left\|q_h^{n+1} - q_h^n\right\|_{L^2}. \end{aligned}$$

As to the last term B_4, we employ (3.35), (3.37) together with the inverse inequality (2.23) to

obtain

$$B_4 \leq C\left(e^{-2\pi^2/k+2\pi\varepsilon/k}+h^{\delta-s}\right)\|q_h^n\|_{\dot{H}^\delta}\left\|q_h^{n+1}-q_h^n\right\|_{\dot{H}^s} \leq Ch^{-s}\|q_h^n\|_{\dot{H}^s}\left\|q_h^{n+1}-q_h^n\right\|_{L^2}.$$

Collecting all four terms to (4.7), canceling the common term $\left\|q_h^{n+1}-q_h^n\right\|_{L^2}$, multiplying by τ, and dropping the higher order terms of h, we get

$$\left\|q_h^{n+1}-q_h^n\right\|_{L^2} \leq Ch^{-s}\tau(\|\boldsymbol{u}^n\|_{\mathbf{L}^\infty}+1)\|q_h^n\|_{\dot{H}^s}+\tau\|f^n\|_{L^2}. \tag{4.8}$$

Inserting (4.8) into (4.6), and applying the definition $\left\|L^{s/2}q_h^n\right\|_{L^2}^2=\|q_h^n\|_{\dot{H}^s}^2$ yield

$$\begin{aligned}\left\|q_h^{n+1}\right\|_{L^2}^2-\|q_h^n\|_{L^2}^2+2\tau\|q_h^n\|_{\dot{H}^s}^2 &\leq \left\|q_h^{n+1}-q_h^n\right\|_{L^2}^2+2\tau|(f^n,q_h^n)|+2\tau\left|(a_h^k-a)(q_h^n,q_h^n)\right| \\ &\leq Ch^{-2s}\tau^2(\|\boldsymbol{u}^n\|_{\mathbf{L}^\infty}+1)^2\|q_h^n\|_{\dot{H}^s}^2+2\tau^2\|f^n\|_{L^2}^2+2\tau\Big(\frac{1}{2}\|f^n\|_{\dot{H}^{-s}}^2+\frac{1}{2}\|q_h^n\|_{\dot{H}^s}^2\Big) \\ &\quad +2\tau\Big(C_IC(s,\varepsilon,\delta)e^{-2\pi^2/k+2\pi\varepsilon/k}h^{s-\delta}\Big)\|q_h^n\|_{\dot{H}^s}^2,\end{aligned}$$

where in the last step we applied $(f^n,q_h^n)\leq\|f^n\|_{\dot{H}^{-s}}\|q_h^n\|_{\dot{H}^s}$ together with a Young's inequality. A reformulation of the above formula immediately gives

$$\begin{aligned}\left\|q_h^{n+1}\right\|_{L^2}^2-\|q_h^n\|_{L^2}^2+\tau\left[1-Ch^{-2s}\tau(\|\boldsymbol{u}^n\|_{\mathbf{L}^\infty}+1)^2-2C_IC(s,\varepsilon,\delta)e^{-2\pi^2/k+2\pi\varepsilon/k}h^{s-\delta}\right]\|q_h^n\|_{\dot{H}^s}^2 \\ \leq C(\tau+1)^2\|f^n\|_{L^2}^2.\end{aligned}$$

Suppose that $h^{-2s}\tau$ and the sinc quadrature step k are taken sufficiently small allowing

$$Ch^{-2s}\tau\left(\|\boldsymbol{u}^n\|_{\mathbf{L}^\infty}+1\right)^2+2C_IC(s,\varepsilon,\delta)e^{-2\pi^2/k+2\pi\varepsilon/k}h^{s-\delta}<c_0<1,$$

for all $0\leq n\leq M$, where $M=T_0/\tau$ and c_0 is a fixed real number. Then for each step $n=$

$0, 1, \cdots, M-1$, we have

$$\left\|q_h^{n+1}\right\|_{L^2}^2 - \left\|q_h^n\right\|_{L^2}^2 + c_0\tau \left\|q_h^n\right\|_{\dot{H}^s}^2 \leq C\tau^2 \left\|f^n\right\|_{L^2}^2 .$$

Therefore, after summing up all terms, we get

$$\max_{0\leq n\leq M-1} \left\|q_h^{n+1}\right\|_{L^2}^2 + c_0 \sum_{n=0}^{M-1} \tau \left\|q_h^n\right\|_{\dot{H}^s}^2 \leq \left\|q_h^0\right\|_{L^2}^2 + C \sum_{n=0}^{M-1} \tau^2 \left\|f^n\right\|_{L^2}^2 .$$

This completes the proof of the lemma. □

4.2 Error estimates for the homogeneous problem

In this section we consider the homogeneous parabolic equation, namely set $f = 0$. It follows from Proposition 5.3.2 of [44] that the solution to (4.1) admits a Dunford-Taylor integral representation

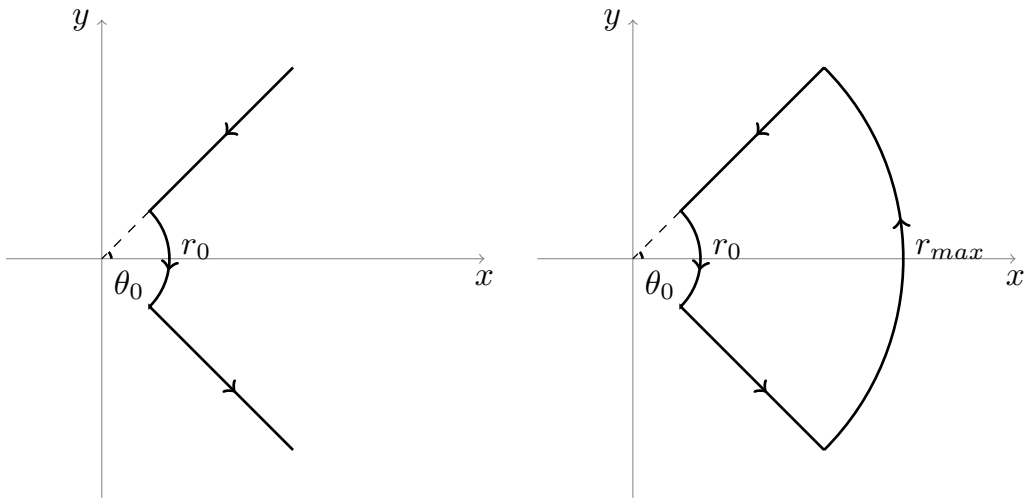
$$q(x;t) = e^{-t\mathcal{A}}v(x) := \frac{1}{2\pi i}\int_{\mathcal{C}} e^{-tz}(zI-\mathcal{A})^{-1}v dz. \tag{4.9}$$

Here the curve $\mathcal{C} = \mathcal{C}(\theta_0, r_0)$ runs in the resolvent set of $\mathcal{A}$. To be specific, $\theta_0 \in (\omega, \pi/2)$, $r_0 \in (0, 1/2)$, and $\mathcal{C}$ is defined by $\mathcal{C}(\theta_0, r_0) = -\mathcal{C}^{(1)}(\theta_0, r_0) - \mathcal{C}^{(2)}(\theta_0, r_0) + \mathcal{C}^{(3)}(\theta_0, r_0)$, where $\mathcal{C}^{(1)}(\theta_0, r_0)$ and $\mathcal{C}^{(3)}(\theta_0, r_0)$ are half lines parametrized by $z = re^{i\theta_0}$ and $z = re^{-i\theta_0}$ with $r \geq r_0$, and $\mathcal{C}^{(2)}(\theta_0, r_0)$ is the arc of circle parametrized by $z = r_0 e^{i\theta}$ with $-\theta_0 \leq \theta \leq \theta_0$ (see Figure 4.1).

As the spectrum set of $\mathcal{A}_h$ is bounded, there exists a closed contour $\mathcal{C}_{r_{max}}$ enclosing the spectrum set, see also Figure 4.1 for example. Here r_{max} is dependent on h. Therefore, the solution to the discrete problem (4.4) could be analogously represented by Dunford-Taylor integral.

$$\begin{aligned} q_h(x;t) = e^{-t\mathcal{A}_h}\pi_h v(x) &:= \frac{1}{2\pi i}\int_{\mathcal{C}} e^{-tz}(zI-\mathcal{A}_h)^{-1}\pi_h v dz \\ &= \frac{1}{2\pi i}\int_{\mathcal{C}_{r_{max}}} e^{-tz}(zI-\mathcal{A}_h)^{-1}\pi_h v \, dz. \end{aligned} \tag{4.10}$$

Before deriving the finite element approximation error we first mention some existing results that are crucial in the remaining of this section. We first define the norm on $D(\mathcal{A}^r)$ for $r \in [0, 1)$



(a) (b)

Figure 4.1. The contours for Dunford-Taylor integral.

by

$$\|v\|_{D(\mathcal{A}^r)} := \|\mathcal{A}^r v\| . \tag{4.11}$$

In addition, we can characterize the domain $D(\mathcal{A}^r)$ for $r \in [0,1]$ following Proposition 3.1,

$$D(\mathcal{A}^r) = \dot{H}^{2rs^*}(\Omega) \tag{4.12}$$

with equivalent norms. Corollary 2.1.8 of [44] implies that for $r \in [0,1]$

$$\|\mathcal{A}^r v\| \leq C \|\mathcal{A}v\|^r \|v\|^{1-r}, \qquad \text{for all } v \in D(\mathcal{A}), \tag{4.13}$$

with the constant C independent of r. We then state the following integration bound which is proved in formula (3.14) in the proof for Theorem 3.1 in [29]. Suppose $a \in (0,1]$ and $t\,, r \geq 0$,

then

$$\int_{\mathcal{C}} \left|e^{-tz}\right| |z|^{a-r-1} \, d|z| \leq \begin{cases} C, & \text{if } a < \min(r,1), \\ C\max\{1, \ln(1/t)\}, & \text{if } r = a < 1, \text{ or } r \geq a = 1, \\ Ct^{-(a-r)}, & \text{if } a > r \geq 0, \end{cases} \tag{4.14}$$

here the constant C does not depend on $z \in \mathcal{C}$, t, a, or r.

We lastly update Lemma 6.3 of [25] in the complex setting.

Lemma 4.2. For $r \in [0,1]$ and $f \in L^2(\Omega)$, there exists a constant C not depending on $z \in \mathcal{C}$ such that

$$\left\|\mathcal{A}^r (zI - \mathcal{A})^{-1} f\right\|_{L^2} \leq C |z|^{r-1} \|f\|_{L^2}. \tag{4.15}$$

The same inequality holds on V_h with $\mathcal{A}$ replaced by $\mathcal{A}_h$, with the constant C being independent of h.

Proof. As $z \in \mathcal{C}$ is in the resolvent set, $(zI - \mathcal{A})^{-1} f$ is in $D(\mathcal{A})$. Applying (4.13) yields

$$\left\|\mathcal{A}^r (zI - \mathcal{A})^{-1} f\right\| \leq C \left\|\mathcal{A}(zI - \mathcal{A})^{-1} f\right\|^r \left\|(zI - \mathcal{A})^{-1} f\right\|^{1-r}.$$

We now estimate the two terms in the right hand side separately. Theorem 2.2 of [41] implies that for $z \in \mathcal{C}^{(1)}$,

$$\left\|(zI - \mathcal{A})^{-1} f\right\| \leq (\sin(\theta_0 - \omega))^{-1} |z|^{-1} \|f\|,$$

while for $z \in \mathcal{C}^{(3)}$,

$$\left\|(zI - \mathcal{A})^{-1} f\right\| \leq |z|^{-1} \|f\|.$$

In addition, Remark 2.1 of [25] ensures that for $z \in \mathcal{C}^{(2)}$, since $\Re(z) \leq |z| \leq r_0 < 1/2$,

$$\left\|(zI - \mathcal{A})^{-1} f\right\| \leq C |z|^{-1} \|f\|.$$

The three estimates immediately implies that for any $z \in \mathcal{C}$,

$$\left\|(zI - \mathcal{A})^{-1} f\right\| \leq C\,|z|^{-1}\,\|f\|\,.$$

In addition, for the second term, it follows that

$$\left\|\mathcal{A}(zI - \mathcal{A})^{-1} f\right\| = \left\|(I - z(zI - \mathcal{A})^{-1}) f\right\| \leq (1 + C)\,\|f\|\,.$$

Collecting the above two inequalities completes the proof. □

The following lemma provides an estimate of of the finite element approximation error.

Lemma 4.3 (Finite element error)**.** Let $s \in [1/2, 1)$ and $\alpha \in (0, 1]$ such that Assumption 2.2 holds. Let $\{\mathcal{T}_h\}$ be a sequence of shape-regular subdivisions of Ω. Let k be the sinc quadrature stepping. For a given h suppose k is chosen to be sufficiently small such that (3.37) holds. Let $\varepsilon \ll 1$ be the constant in Lemma 3.5. Suppose $q(t)$ and $q_h(t)$ are solutions to (4.1) and (4.4) respectively with $f = 0$. Assume the initial condition $v \in \mathbb{H}^{2\gamma}$ with $\gamma \in [0, (1 + \alpha)/2]$, then for all $t > 0$,

$$\|q(\cdot, t) - q_h(\cdot, t)\|_{L^2} \leq D(t) \ln(1/h) \Big(e^{-2\pi^2/k + 2\pi\varepsilon/k} + \ln(1/h) h^{2\alpha} \Big) \|v\|_{\mathbb{H}^{2\gamma}}\,, \tag{4.16}$$

with

$$D(t) := \begin{cases} C, & \text{if } \gamma > \alpha^*, \\ C \max\{1, \ln(1/t)\}, & \text{if } \gamma = \alpha^*, \\ C t^{-(\alpha^* - \gamma)/s^*}, & \text{if } \gamma < \alpha^*, \end{cases} \tag{4.17}$$

where $\alpha^* = \alpha$ if $s = s^*$, and $\alpha^* = \min(\alpha + \rho/2, (1 + \alpha)/2, 2s^*)$ if $s > s^*$, with ρ the constant in Proposition 3.1.

Proof. We first apply a triangle inequality to obtain

$$\|q - q_h\| \leq \|q - \pi_h q\| + \|\pi_h q - q_h\| =: I + II.$$

The two terms I and II are then estimated separately.

1 When $\gamma \geq \alpha^*$, the L^2-projection estimate (2.20) implies

$$I \leq Ch^{2\alpha} \|q\|_{\dot{H}^{2\alpha}} \leq Ch^{2\alpha} \left\|e^{-t\mathcal{A}}v\right\|_{\dot{H}^{2\alpha^*}} \leq Ch^{2\alpha} \|v\|_{\dot{H}^{2\alpha^*}} \leq Ch^{2\alpha} \|v\|_{\dot{H}^{2\gamma}} .$$

When $\gamma < \alpha^*$, following the property of semi-group,

$$\begin{aligned} I &\leq Ch^{2\alpha} \|q\|_{\dot{H}^{2\alpha}} \leq Ch^{2\alpha} \left\|e^{-t\mathcal{A}}v\right\|_{\dot{H}^{2\alpha^*}} \\ &\leq Ch^{2\alpha} \left\|\mathcal{A}^{(\alpha^*-\gamma)/s^*} e^{-t\mathcal{A}}v\right\|_{\dot{H}^{2\gamma}} \leq Ch^{2\alpha} t^{-(\alpha^*-\gamma)/s^*} \|v\|_{\dot{H}^{2\gamma}} . \end{aligned}$$

2 As to II, we start with the Dunford-Taylor integral formulas (4.9) and (4.10),

$$\begin{aligned} II = \|\pi_h q - q_h\| &= \left\| \frac{1}{2\pi i} \int_{\mathcal{C}} e^{-tz} \left[\pi_h (zI - \mathcal{A})^{-1} - (zI - \mathcal{A}_h)^{-1}\pi_h\right] v dz \right\| \\ &\leq \frac{1}{2\pi} \int_{\mathcal{C}} \left|e^{-tz}\right| \underbrace{\left\| \left[\pi_h (zI - \mathcal{A})^{-1} - (zI - \mathcal{A}_h)^{-1}\pi_h\right] v \right\|}_{=:G(z)} d\,|z| , \end{aligned}$$

where again the last inequality applied Minkowski's inequality of integral form. Notice that

$$\begin{aligned} G(z) &= \left\| \left[\pi_h \mathcal{A}^{-1}(z\mathcal{A}^{-1} - I)^{-1} - (z\mathcal{A}_h^{-1} - I)^{-1}\pi_h \mathcal{A}_h^{-1}\right] v \right\| \\ &= \left\| (z\mathcal{A}_h^{-1} - I)^{-1}\pi_h (\mathcal{A}^{-1} - \mathcal{A}_h^{-1})(z\mathcal{A}^{-1} - I)^{-1} v \right\| \\ &\leq \left\| (z\mathcal{A}_h^{-1} - I)^{-1}\pi_h \right\|_{\dot{H}^{-\alpha^*+s^*+\eta} \to L^2} \left\| \mathcal{A}^{-1} - \mathcal{A}_h^{-1} \right\|_{\dot{H}^{\alpha^*-s^*+\eta} \to \dot{H}^{-\alpha^*+s^*+\eta}} \\ &\quad \left\| (z\mathcal{A}^{-1} - I)^{-1} \right\|_{D(\mathcal{A}^{\gamma/s^*}) \to \dot{H}^{\alpha^*-s^*+\eta}} \|v\|_{D(\mathcal{A}^{\gamma/s^*})} , \end{aligned}$$

where $\eta \in [\alpha^* - s^*, \alpha^* + s^*]$. Directly from Proposition 3.1 and Theorem 3.13 (let $\beta = 2\alpha^*$), we obtain that

$$\left\| (\mathcal{A}^{-1} - \mathcal{A}_h^{-1}) \right\|_{\dot{H}^{\alpha^*-s^*+\eta} \to \dot{H}^{-\alpha^*+s^*+\eta}} \leq C \ln(1/h) \Big(e^{-2\pi^2/k + 2\pi\varepsilon/k} + \ln(1/h) h^{2\alpha} \Big).$$

In addition, the characterization (4.12) yields

$$\|v\|_{D(\mathcal{A}^{\gamma/s^*})} \leq C \|v\|_{\mathbb{H}^{2\gamma}}.$$

We then distinguish different cases for the remaining two terms.

3 Case 1: $\gamma \in [\max(0, \alpha^* - s^*), \min(\alpha^*, s^*)]$. We let $\eta = -s^* + \alpha^*$. First observe that from the stability of L^2-projection and the discrete version of inequality (4.15),

$$\left\|(z\mathcal{A}_h^{-1} - I)^{-1}\pi_h\right\|_{L^2\to L^2} \leq \left\|\mathcal{A}_h(zI - \mathcal{A}_h)^{-1}\right\|_{L^2\to L^2} \|\pi_h\|_{L^2\to L^2} \leq C.$$

Additionally, we obtain from (4.15) that since $(\alpha^* - \gamma)/s^* \in [0, 1]$,

$$\left\|(z\mathcal{A}^{-1} - I)^{-1}\right\|_{D(\mathcal{A}^{\gamma/s^*})\to\dot{H}^{2\alpha^*-2s^*}} \leq C \left\|\mathcal{A}^{(\alpha^*-\gamma)/s^*}(zI - \mathcal{A})^{-1}\right\|_{L^2\to L^2} \leq C|z|^{(\alpha^*-\gamma)/s^*-1}.$$

Case 2: $\gamma \in [\alpha^*, (1 + \alpha)/2] \subset [\alpha^*, \alpha^* + s^*]$. We choose $\eta = s^* + \alpha^*$. Then

$$\left\|(z\mathcal{A}_h^{-1} - I)^{-1}\pi_h\right\|_{\dot{H}^{2s^*}\to L^2} \leq \left\|(zI - \mathcal{A}_h)^{-1}\right\|_{L^2\to L^2} \|\pi_h\|_{\dot{H}^{2s^*}\to\dot{H}_h^{2s^*}} \leq C|z|^{-1}.$$

While since $(\alpha^* - \gamma)/s^* \in [-1, 0]$,

$$\left\|(z\mathcal{A}^{-1} - I)^{-1}\right\|_{D(\mathcal{A}^{\gamma/s^*})\to\dot{H}^{2\alpha^*}} \leq C \left\|\mathcal{A}^{(\alpha^*-\gamma)/s^*+1}(zI - \mathcal{A})^{-1}\right\|_{L^2\to L^2} \leq C|z|^{(\alpha^*-\gamma)/s^*}.$$

Case 3: $\gamma \in (\min(\alpha^*, s^*), \alpha^*)$. In this case $\alpha^* > s^*$ and we pick $\eta = \gamma$. Then since $(\alpha^* - \gamma)/(2s^*) \in [0, 1/2]$, we obtain

$$\begin{aligned}\left\|(z\mathcal{A}_h^{-1} - I)^{-1}\pi_h\right\|_{\dot{H}^{\gamma-\alpha^*+s^*}\to L^2} &\leq \left\|\mathcal{A}_h^{(\alpha^*-\gamma+s^*)/(2s^*)}(zI - \mathcal{A}_h)^{-1}\right\|_{L^2\to L^2} \|\pi_h\|_{\dot{H}^{\gamma-\alpha^*+s^*}\to\dot{H}_h^{\gamma-\alpha^*+s^*}} \\ &\leq C|z|^{(\alpha^*-\gamma)/(2s^*)-1/2},\end{aligned}$$

and

$$\begin{aligned}\left\|(z\mathcal{A}^{-1}-I)^{-1}\right\|_{D(\mathcal{A}^{\gamma/s^*})\to\dot{H}^{\alpha^*+\gamma-s^*}} &\le C\left\|\mathcal{A}^{(\alpha^*-\gamma+s^*)/(2s^*)}(zI-\mathcal{A})^{-1}\right\|_{L^2\to L^2}\\ &\le C\,|z|^{(\alpha^*-\gamma)/(2s^*)-1/2}.\end{aligned}$$

Gathering all above three cases altogether, we obtain

$$G(z)\le C\ln(1/h)\Big(e^{-2\pi^2/k+2\pi\varepsilon/k}+\ln(1/h)h^{2\alpha}\Big)\,|z|^{(\alpha^*-\gamma)/s^*-1}\,\|v\|_{\mathbb{H}^{2\gamma}}.$$

What remains is to show that

$$\int_{\mathcal{C}}\left|e^{-tz}\right|\,|z|^{(\alpha^*-\gamma)/s^*-1}\;d\,|z|\le D(t),$$

which is a direct consequence of (4.14).

4 Case 4: $\gamma\in[0,\max(0,\alpha^*-s^*))$. In this case $\alpha^*>s^*$ and we pick $\eta=\gamma-s^*$. Then notice that $(\alpha^*-\gamma)/(2s^*)\in[1/2,1]$, inequality (4.15), the inverse inequality (2.23), the π_h stability, together with the time stepping stability condition $\tau\le Ch^{2s}$ yield

$$\begin{aligned}\left\|(z\mathcal{A}_h^{-1}-I)^{-1}\pi_h\right\|_{\dot{H}^{\gamma-\alpha^*}\to L^2} &\le \left\|(z\mathcal{A}_h^{-1}-I)^{-1}\right\|_{\dot{H}_h^{\gamma-\alpha^*+2s^*}\to L^2}\|\pi_h\|_{\dot{H}^{\gamma-\alpha^*}\to\dot{H}_h^{\gamma-\alpha^*+2s^*}}\\ &\le Ch^{-2s^*}\left\|\mathcal{A}_h^{(\alpha^*-\gamma)/(2s^*)}(zI-\mathcal{A}_h)^{-1}\right\|_{L^2\to L^2}\le C\tau^{-1}\,|z|^{(\alpha^*-\gamma)/(2s^*)-1},\end{aligned}$$

and

$$\left\|(z\mathcal{A}^{-1}-I)^{-1}\right\|_{D(\mathcal{A}^{\gamma/s^*})\to\dot{H}^{\alpha^*+\gamma-2s^*}}\le C\left\|\mathcal{A}^{(\alpha^*-\gamma)/(2s^*)}(zI-\mathcal{A})^{-1}\right\|_{L^2\to L^2}\le C\,|z|^{(\alpha^*-\gamma)/(2s^*)-1}.$$

From (4.14) we get

$$\tau^{-1}\int_{\mathcal{C}}\left|e^{-tz}\right|\,|z|^{(\alpha^*-\gamma)/s^*-2}\;d\,|z|\le C\tau^{-1}t^{-(\alpha^*-\gamma)/s^*+1}\le Ct^{-(\alpha^*-\gamma)/s^*},$$

this completes the proof. □

Remark 4.4. The special case $\boldsymbol{u} = \mathbf{0}$ has been thoroughly discussed in [29]. Indeed, the error estimate in Lemma 4.3 is consistent with Corollary 4.1 of [29], where $s^* = s$ and $\alpha^* = \alpha$.

Error estimate for time discretization

Next step we apply the stability result of forward Euler method to estimate the error introduced by time discretization. Recall from (4.5) that the finite element approximation q_h^n satisfies

$$q_h^n = E_\tau q_h^{n-1} + \tau \pi_h f(t_n), \tag{4.18}$$

with $E_\tau := (I - \tau \mathcal{A}_h)$. For $f = 0$ in particular, applying E_τ recursively from t_{n-1} to $t_0 = 0$ yields

$$q_h^n = E_\tau q_h^{n-1} = \cdots = E_\tau^n \pi_h v.$$

Also recall from (4.10) that $q_h(t_n) = E_h(\tau) q_h(t_{n-1}) = \cdots = E_h(t_n)\pi_h v$ with $E_h(\tau) := e^{-\tau \mathcal{A}_h}$. The following lemma provides an error estimate between the two operators E_τ^n and $E_h(t_n)$.

Lemma 4.5. Assume that $\tau r_{max} \le \lambda_0$ for some $0 < \lambda_0 < 2$. For $v_h \in V_h$, there exists a constant C not depending on h, n, or τ, such that

$$\left\| \left(E_\tau^n - E_h(t_n)\right) v_h \right\| \le C \max(1, \ln(1/\tau)) t_n^{\gamma/s^* - 1} \tau \left\| v_h \right\|_{\dot{H}_h^{2\gamma}}, \tag{4.19}$$

where $\gamma \in [0, s^*]$ and $t_n = n\tau$ with n nonnegative integer.

Proof. In view of (4.12), we rewrite (4.19) as

$$\left\| \mathcal{A}_h^{-\gamma/s^*} \left(E_\tau^n - E_h(t_n)\right) v_h \right\| \le C \max(1, \ln(1/\tau)) n^{\gamma/s^* - 1} \tau^{\gamma/s^*} \left\| v_h \right\|.$$

Clearly both $\mathcal{A}_h^{-\gamma/s^*} E_\tau^n = \mathcal{A}_h^{-\gamma/s^*} (I - \tau \mathcal{A}_h)^n$ and $\mathcal{A}_h^{-\gamma/s^*} E_h(t_n) = \mathcal{A}_h^{-\gamma/s^*} e^{-n\tau \mathcal{A}_h}$ are bounded

operators. The two terms can also be represented with Dunford-Taylor integral (4.10) to have

$$\begin{aligned}\left\|\mathcal{A}_h^{-\gamma/s^*}\left(E_\tau^n - E_h(t_n)\right)v_h\right\| &= \left\|\frac{1}{2\pi i}\int_{\mathcal{C}} z^{-\gamma/s^*}(e^{-n\tau z} - (1-\tau z)^n)(zI-\mathcal{A}_h)^{-1}v_h dz\right\| \\ &\le \frac{1}{2\pi}\int_{\mathcal{C}} \left|z^{-\gamma/s^*}(e^{-n\tau z} - (1-\tau z)^n)\right| \left\|(zI-\mathcal{A}_h)^{-1}v_h\right\| d\left|z\right|.\end{aligned}$$

Invoking Taylor expansion, there exist constants $C > 0$ and $c \in (0,1)$ such that for $z \in \mathcal{C}_{r_{max}}$,

$$\left|e^{-\tau z} - (1-\tau z)\right| \le C\tau^2 \left|z\right|^2, \qquad \text{and } \left|1-\tau z\right| < \left|e^{-c\tau z}\right|. \tag{4.20}$$

It follows that

$$\begin{aligned}\left|e^{-n\tau z} - (1-\tau z)^n\right| &= \left|(e^{-\tau z} - (1-\tau z))\sum_{j=0}^{n-1}(1-\tau z)^{n-1-j}e^{-j\tau z}\right| \\ &\le Cn\tau^2\left|z\right|^2 e^{-c(n-1)\tau|z|\cos\theta_0} \le C(\tau\left|z\right|)^{\gamma/s^*}n^{\gamma/s^*-1}.\end{aligned}$$

In addition, (4.15) yields

$$\left\|(zI-\mathcal{A}_h)^{-1}v_h\right\| \le C\left|z\right|^{-1}\left\|v_h\right\|.$$

Collecting the above two estimates we have

$$\begin{aligned}&\int_{\mathcal{C}} \left|z^{-\gamma/s^*}(e^{-n\tau z} - (1-\tau z)^n)\right| \left\|(zI-\mathcal{A}_h)^{-1}v_h\right\| d\left|z\right| \\ &\qquad \le C\tau^{\gamma/s^*}n^{\gamma/s^*-1}\left\|v_h\right\|\int_{\mathcal{C}}\left|z\right|^{-1}d\left|z\right| \\ &\qquad \le C\max(1,\ln(1/\tau))\tau^{\gamma/s^*}n^{\gamma/s^*-1}\left\|v_h\right\|,\end{aligned}$$

where to derive the last inequality we applied (4.14). The proof is complete. □

The time discretization error is the direct consequence of the above lemma.

Lemma 4.6 (Time discretization error)**.** Let $t_n = n\tau$ with $\tau > 0$ the time stepping and n the number of time stepping. Assume $\tau r_{max} \le \lambda_0$ for some $0 < \lambda_0 < 2$. Let q_h^n and $q_h(t)$ given by

(4.5) and (4.4) respectively with $f = 0$. Assume initial data $v \in \dot{H}^{2\gamma}$, $\gamma \in [0, s^*]$ with s^* defined by (3.8). Then

$$\|q_h^n - q_h(t_n)\| \leq C \max(1, \ln(1/\tau)) t_n^{\gamma/s^*-1} \tau \|v\|_{\mathbb{H}^{2\gamma}} . \tag{4.21}$$

We end our discussion for the homogeneous problem by the following theorem, which is derived from applying the triangle inequality to Lemma 4.3 and Lemma 4.6.

Theorem 4.7 (Fully discretization error)**.** Let $s \in [1/2, 1)$ and $\alpha \in (0, 1]$ such that Assumption 2.2 holds. Let $\{\mathcal{T}_h\}$ be a sequence of shape-regular subdivisions of Ω. Let k be the sinc quadrature stepping. For a given h suppose k is chosen to be sufficiently small such that (3.37) holds. Let $\varepsilon \ll 1$ be the constant in Lemma 3.5. Let $t_n = n\tau$ with $\tau > 0$ the time stepping and n the number of time stepping. Assume $\tau r_{max} < \lambda_0$ for some $0 < \lambda_0 < 2$. Suppose $q(t)$ and q_h^n are solutions to (4.1) and (4.5) respectively with $f = 0$. Assume the initial condition $v \in \mathbb{H}^{2\gamma}$ with $\gamma \in [0, (1 + \alpha)/2]$. Then,

$$\|q(t_n) - q_h^n\| \leq \Big(D(t_n) C(h) \big(e^{-2\pi^2/k + 2\pi\varepsilon/k} + h^{2\alpha} \big) + C \max(1, \ln(1/\tau)) t_n^{\gamma/s^*-1} \tau \Big) \|v\|_{\mathbb{H}^{2\gamma}} ,$$

with s^* defined by (3.8), $D(t_n)$ defined by (4.17) and $C(h)$ defined by (3.31).

4.3 Error estimates for the inhomogeneous problem

In this subsection we consider the inhomogeneous problem. Namely, in the weak formulations (4.1) and (4.4) we assume the right hand data $f \in L^\infty(0, \mathtt{T}; \mathbb{H}^{2\gamma})$ for $\gamma \in [0, (1 + \alpha)/2]$, and the initial condition $v = 0$. Under such circumstance, after invoking Duhamel's principle we have that the solution to (4.1) is

$$q(x; t) = \int_0^t e^{-y\mathcal{A}} f(t - y) dy, \tag{4.22}$$

and the solution to (4.4) is

$$q_h(x; t) = \int_0^t e^{-y\mathcal{A}_h} \pi_h f(t - y) dy. \tag{4.23}$$

Combining the two expressions we obtain

$$\begin{aligned}\|q(\cdot,t)-q_h(\cdot,t)\| &\le \left\| \int_0^t \left(e^{-y\mathcal{A}} - e^{-y\mathcal{A}_h}\pi_h\right) f(t-y)dy \right\| \\ &\le \int_0^t \left\| \left(e^{-y\mathcal{A}} - e^{-y\mathcal{A}_h}\pi_h\right) f(t-y) \right\| dy.\end{aligned}$$

When $\gamma > \alpha^*$, we achieve the optimal convergence rate $h^{2\alpha}$ since that from Lemma 4.3,

$$\begin{aligned}\|q(\cdot,t)-q_h(\cdot,t)\| &\le C(h)\left(e^{-2\pi^2/k+2\pi\varepsilon/k} + h^{2\alpha}\right) \|f\|_{L^\infty(0,T_0;\mathbb{H}^{2\gamma})} \int_0^t D(y)\, dy \\ &\le C(h)t\left(e^{-2\pi^2/k+2\pi\varepsilon/k} + h^{2\alpha}\right) \|f\|_{L^\infty(0,T_0;\mathbb{H}^{2\gamma})}.\end{aligned}$$

When $\gamma < \alpha^*$, we introduce an intermediate parameter

$$\bar{\alpha} := \min(\alpha^*, \gamma + s^* - \epsilon), \tag{4.24}$$

for some sufficiently small $\epsilon > 0$. It is trivial to check that in this case

$$\int_0^t D(y)\, dy \le C \int_0^t y^{-(\bar{\alpha}-\gamma)/s^*}\, dy = \frac{C}{1+(\bar{\alpha}-\gamma)/s^*} t^{1-(\bar{\alpha}-\gamma)/s^*} < \infty.$$

In conclusion, we have the following lemma.

Lemma 4.8 (Finite element error)**.** Let $s \in [1/2, 1)$ and $\alpha \in (0, 1]$ such that Assumption 2.2 holds. Let k be the SINC quadrature stepping. Let $\varepsilon \ll 1$ be the constant in Lemma 3.5. Let $\{\mathcal{T}_h\}$ be a sequence of shape-regular subdivisions of Ω. Suppose $q(t)$ and $q_h(t)$ are solutions to (4.1) and (4.4) with $v = 0$ respectively, and suppose the right hand data $f \in L^\infty(0, \mathtt{T}\,; \mathbb{H}^{2\gamma})$ with $\gamma \in [0, (1+\alpha)/2]$. Then for all $t > 0$,

$$\|q(\cdot,t)-q_h(\cdot,t)\|_{L^2(\Omega)} \le \tilde{D}(t)C(h)\left(e^{-2\pi^2/k+2\pi\varepsilon/k} + h^{2\bar{\alpha}}\right) \|f\|_{L^\infty(0,T_0;\mathbb{H}^{2\gamma})}. \tag{4.25}$$

with $\bar{\alpha}$ defined by (4.24), $C(h)$ defined by (3.31) and

$$\tilde{D}(t) = \begin{cases} Ct, & \text{if } \gamma > \alpha^*, \\ Ct\max(1, \ln(1/t)), & \text{if } \gamma = \alpha^*, \\ Ct^{1-(\bar{\alpha}-\gamma)/s^*}, & \text{if } \gamma < \alpha^*. \end{cases} \tag{4.26}$$

Next we continue with analyzing the error introduced by forward Euler time stepping.

Lemma 4.9 (Time discretization error)**.** Let $s \in [1/2, 1)$. Let $\alpha \in (0, 1]$ such that Assumption 2.2 holds. Let $t_n = n\tau$ with $\tau > 0$ and n nonnegative integer. Assume $\tau r_{max} < \lambda_0$ for some $0 < \lambda_0 < 2$. Suppose $q_h(t)$ and q_h^n are solutions to (4.4) and (4.5) with $v = 0$ respectively. Assume the right hand data satisfies $f \in L^\infty(0, t_n; \mathbb{H}^{2\gamma})$ and $f_t \in L^1(0, t_n; L^2)$ with $\gamma \in [0, \min(s^*, (1+\alpha)/2)]$. Then

$$\|q_h^n - q_h(t_n)\| \le C\tau \left(\max(1, \ln(1/\tau)) t_n^{\gamma/s^*} \|f\|_{L^\infty(0,t_n;\mathbb{H}^{2\gamma})} + \|f_t\|_{L^1(0,t_n;L^2)} \right),$$

with constant C not depending on h or τ.

Proof. 1 Recall from the recursive formula (4.18) that

$$q_h^n = \tau \sum_{j=0}^{n-1} E_\tau^{n-1-j} \pi_h f(t_j).$$

As to $q_h(t_n)$, we invoke a change of variable $y = t_j + \tau\xi$ for each subinterval $[t_j, t_{j+1}]$ for $j = 0, \cdots, n-1$ to rewrite (4.23) as

$$\begin{aligned} q_h(t_n) &= \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} e^{-(t_n - y)\mathcal{A}_h} \pi_h f(y) dy = \tau \sum_{j=0}^{n-1} e^{-(n-1-j)\tau\mathcal{A}_h} \int_0^1 e^{-(\tau - \tau\xi)\mathcal{A}_h} \pi_h f(t_j + \tau\xi) d\xi \\ &= \tau \sum_{j=0}^{n-1} E_h(t_{n-1-j}) \int_0^1 E_h(\tau - \tau\xi) \pi_h f(t_j + \tau\xi) d\xi =: \tau \sum_{j=0}^{n-1} E_h(t_{n-1-j}) I_\tau f(t_j). \end{aligned}$$

Here $E_h(\tau) := e^{-\tau\mathcal{A}_h}$, and $I_\tau f(t_j) := \int_0^1 E_h(\tau - \tau\xi)\pi_h f(t_j + \tau\xi) d\xi$. With the above expressions

the error $e_h^n := q_h^n - q_h(t_n)$ can be represented as

$$\begin{aligned} e_h^n &= \tau \sum_{j=0}^{n-1} \left(E_\tau^{n-1-j} \pi_h f(t_j) - E_h(t_{n-1-j}) I_\tau f(t_j) \right) \\ &= \tau \sum_{j=0}^{n-1} \left(E_\tau^{n-1-j} - E_h(t_{n-1-j}) \right) \pi_h f(t_j) + \tau \sum_{j=0}^{n-1} E_h(t_{n-1-j}) \left(\pi_h f(t_j) - I_\tau f(t_j) \right) \\ &=: e_{h,1}^n + e_{h,2}^n. \end{aligned}$$

We then estimate the two errors separately.

2 For $e_{h,1}^n$, we invoke Lemma 4.5 for each term in the summand, which yields

$$\begin{aligned} \left\| e_{h,1}^n \right\| &\le C \max(1, \ln(1/\tau)) \tau^2 \sum_{j=0}^{n-1} t_{n-j-1}^{\gamma/s^*-1} \left\| f(t_j) \right\|_{\dot{H}^{2\gamma}} \\ &\le C \max(1, \ln(1/\tau)) \tau \left\| f \right\|_{L^\infty(0,t_n;\dot{H}^{2\gamma})} \int_0^{t_n} \xi^{\gamma/s^*-1} \, d\xi \\ &\le C \max(1, \ln(1/\tau)) \tau t_n^{\gamma/s^*} \left\| f \right\|_{L^\infty(0,t_n;\dot{H}^{2\gamma})}. \end{aligned} \tag{4.27}$$

3 In order to estimate $e_{h,2}^n$, we first employ Taylor expansion of $f(t_j + \tau\xi)$ at t_j with integral remainder to rewrite $I_\tau f(t_j)$ as

$$I_\tau f(t_j) = \Big(\int_0^1 E_h(\tau - \tau\xi) \, d\xi \Big) \pi_h f(t_j) + \int_0^1 E_h(\tau - \tau\xi) \pi_h \Big(\int_{t_j}^{t_j + \tau\xi} f_t(\eta) \, d\eta \Big) \, d\xi.$$

Because the operator $E_h((1-\xi)\tau)$ is bounded for $\xi \in (0,1)$, the $L^2(\Omega)$-norm of the second term on the right hand side is bounded by $C \int_{t_j}^{t_{j+1}} \left\| f_t(\eta) \right\| d\eta$. In order to obtain error bound for each term in the finite sum in $e_{h,2}^n$, we first see that

$$\begin{aligned} &\left\| E_h(t_{n-1-j}) \Big(I - \int_0^1 E_h(\tau - \tau\xi) d\xi \Big) \pi_h f(t_j) \right\| \\ &\quad \le C \left\| \mathcal{A}_h^{-\gamma/s^*} E_h(t_{n-1-j}) \Big(I - \int_0^1 E_h(\tau - \tau\xi) d\xi \Big) \right\|_{L^2 \to L^2} \left\| \pi_h f(t_j) \right\|_{\dot{H}_h^{2\gamma}}. \end{aligned}$$

While for $v_h \in V_h$,

$$\begin{aligned}
&\left\| \mathcal{A}_h^{-\gamma/s^*} E_h(t_{n-1-j}) \Big(I - \int_0^1 E_h(\tau - \tau\xi) d\xi \Big) v_h \right\| \\
&\quad \le \frac{1}{2\pi} \int_{\mathcal{C}} |z|^{-\gamma/s^*} \left| e^{-t_{n-j-1}z} \right| \underbrace{\left| 1 - \int_0^1 e^{-(1-\xi)\tau z} d\xi \right|}_{=:B_1} \underbrace{\left\| (zI - \mathcal{A}_h)^{-1} v_h \right\|}_{=:B_2} \, d\,|z| \\
&\quad \le C \int_{\mathcal{C}} |z|^{-\gamma/s^*} \left| e^{-t_{n-j-1}z} \right| |\tau z| \, |z|^{-1} \, \|v_h\| \, \, d\,|z| \\
&\quad \le C\tau t_{n-j-1}^{\gamma/s^*-1} \, \|v_h\| \,,
\end{aligned}$$

where for the second inequality we applied (4.20) for term B_1, and the discrete version of (4.15) with $r = 0$ for term B_2, while for the last inequality we applied (4.14) with $r = 0$, and $a = 1-\gamma/s^*$. Combining the two estimates and applying a triangle inequality yield for each $j = 0, \cdots, n-1$,

$$\|E_h(t_{n-1-j}) \left(\pi_h f(t_j) - I_\tau f(t_j) \right)\| \le C \Big(\int_{t_j}^{t_{j+1}} \|f_t(\eta)\| \, d\eta + \tau t_{n-j-1}^{\gamma/s^*-1} \, \|f\|_{L^\infty(0,t_n;\dot{H}^{2\gamma})} \Big).$$

Collecting all the n terms yield the error bound for $e_{h,2}^n$. Namely,

$$\begin{aligned}
\left\| e_{h,2}^n \right\| &\le C\tau \sum_{j=0}^{n-1} \Big(\int_{t_j}^{t_{j+1}} \|f_t(\eta)\| \, d\eta + \tau t_{n-j-1}^{\gamma/s^*-1} \, \|f\|_{L^\infty(0,t_n;\dot{H}^{2\gamma})} \Big) \\
&\le C\tau \Big(t_n^{\gamma/s^*} \, \|f\|_{L^\infty(0,t_n;\dot{H}^{2\gamma})} + \|f_t\|_{L^1(0,t_n;L^2)} \Big).
\end{aligned} \tag{4.28}$$

4 The proof is complete after combining the two error estimates (4.27) and (4.28), and invoking the norm equivalence inequalities (2.7) and (2.16). □

Combining the results Lemma 4.8 and Lemma 4.9, we end this section by the following full discretization error estimate.

Theorem 4.10 (Full discretization error)**.** Let $s \in [1/2, 1)$, and $\alpha \in (0, 1]$ such that Assumption 2.2 holds. Let k be the sinc quadrature spacing. Let $\varepsilon \ll 1$ be the constant in Lemma 3.5. Let $\{\mathcal{T}_h\}$ be a sequence of shape-regular subdivisions of Ω. Let $t_n = n\tau$ with $\tau > 0$ and n nonnegative

integer. Assume $\tau r_{max} < \lambda_0$ for some $0 < \lambda_0 < 2$. Suppose $q(t)$ and q_h^n are solutions to (4.1) and (4.5) with $v = 0$ respectively. Assume the right hand data satisfies $f \in L^\infty(0, t_n; \mathbb{H}^{2\gamma})$ and $f_t \in L^1(0, t_n; L^2)$ with $\gamma \in [0, (1+\alpha)/2]$. Then

$$\begin{aligned}\|q(t_n) - q_h^n\| \leq & C\tau \left(\max(1, \ln(1/\tau)) t_n^{\gamma/s^*} \|f\|_{L^\infty(0,t_n;\mathbb{H}^{2\gamma})} + \|f_t\|_{L^1(0,t_n;L^2)} \right) \\ & + \tilde{D}(t) C(h) \left(e^{-2\pi^2/k + 2\pi\varepsilon/k} + h^{2\bar{\alpha}} \right) \|f\|_{L^\infty(0,T_0;\mathbb{H}^{2\gamma})},\end{aligned}$$

with constant C not depending on k, h or τ, and parameters s^* defined by (3.8), $\bar{\alpha}$ defined by (4.24), $\tilde{D}(t)$ defined by (4.26) and $C(h)$ defined by (3.31).

4.4 Numerical examples

Consider the inhomogeneous problem in the square domain $\Omega = (0,1)^2$. Let the solution

$$q(t, x, y) = t^2 \sin(m\pi x) \sin(n\pi y)$$

and the velocity $\boldsymbol{u} = (U_1, U_2)$ such that the right hand side data admits

$$\begin{aligned}f = & (2t + ((m^2+n^2)\pi^2)^s t^2) \sin(m\pi x) \sin(n\pi y) \\ & + \pi t^2 (U_1 m \cos(m\pi x) \sin(n\pi y) + U_2 n \sin(m\pi x) \cos(n\pi y)).\end{aligned}$$

We set $m = 1$, $n = 2$, $U_1 = U_2 = 1$. We fix the time step $\tau = 10^{-4}$ and sinc quadrature step $k = 0.2$ so that the FEM error dominates. We vary the mesh sizes $h = 1/2^j$, $j = 5, 6, 7$ for different values of s. The L^2 error between the approximated solution q_h^n and exact solution q at time $\mathrm{T} = 0.5$ is reported in Table 4.1. In all cases, the optimal convergence rate h^2 are observed.

In addition, we examine the convergence rate of time discretization by fixing the mesh size $h = 1/32$ (refine 4 times from coarse mesh) and the sinc quadrature size $k = 0.2$. Due to the stability requirement, the time stepping error is difficult to dominate. Alternatively, we first compute the error for time stepping $\tau = 10^{-5}$ and use it as the approximated space discretization error. We then

	$s = 0.5$		$s = 0.6$		$s = 0.7$		$s = 0.8$		$s = 0.9$	
h	L^2 err	ORC	L^2 err	ORC	L^2 err	ORC	L^2 err	ORC	L^2 err	ORC
1/32	2.10E-4		2.30E-4		2.59E-4		2.93E-4		3.29E-4	
1/64	5.43E-5	1.95	5.94E-5	1.96	6.63E-5	1.97	7.44E-5	1.98	8.23E-5	2.00
1/128	1.40E-5	1.95	1.52E-5	1.96	1.69E-5	1.97	1.88E-5	1.98	2.06E-5	2.00

Table 4.1. L^2 error at $\mathtt{T} = 0.5$ for space discretization and observed rate of convergence (ORC) for the inhomogeneous problem for different values of s. According to Lemma 4.8, the theoretical convergence rate for space discretization is 2 for all cases.

compute for $\tau = 0.004, 0.002, 0.001$ and subtract the approximated space discretization error from the total error. In correspondence to the stability condition $\tau \leq Ch^{2s}$, smaller time stepping are used for larger s. Table 4.2 shows the asymptotic observed convergence rate at $\mathtt{T} = 0.5$.

	$s = 0.5$		$s = 0.6$		$s = 0.7$		$s = 0.8$		$s = 0.9$	
τ	L^2 err	ORC	L^2 err	ORC	L^2 err	ORC	L^2 err	ORC	L^2 err	ORC
0.002	5.30E-5		4.99E-5		4.20E-5		3.57E-5		5.93E-6	
0.001	2.50E-5	1.08	2.42E-5	1.04	2.07E-5	1.02	1.61E-5	1.15	4.32E-6	0.46
0.0005	1.20E-5	1.06	1.18E-5	1.03	1.02E-5	1.02	7.95E-6	1.02	2.15E-6	1.01

Table 4.2. L^2 error at $\mathtt{T} = 0.5$ for time discretization and observed rate of convergence (ORC) for the inhomogeneous problem for different values of s. According to Lemma 4.9, the theoretical convergence rate for time discretization is 1 for all cases.

5. APPLICATION TO THE SURFACE QUASI-GEOSTROPHIC SYSTEM

In the 1940s Charney [45] introduced the quasi-geostrophic model which describes the large-scale atmospheric motions. The model essentially relies on the conservation of potential vorticity in the interior of geostrophic flow. The approximation straightifies the geometric curvature on the motion (which is of minor influence except for extremely large-scale motions), but not the variation with the latitude on the Coriolis force. In fact, the Coriolis force is essential for the explanation of the point-wise dynamics of the motion; as illustrated in Section 5.1. The simplification in the domain is adopted in this work and without losing of generality we define $\Omega := (0, 2\pi L)^2$ with $L > 0$ the horizontal length scale, and denote $\mathcal{C} := \Omega \times (0, \infty)$.

The greatly-developed quasi-geostrophic theory has been proven successful in describing major features of atmospheric motions of horizontal scales between 10 and 500 km [31, 32, 33], as well as the oceanic frontal structures of scales a few tens of kilometers [46, 47, 48]. In particular, the surface quasi-geostrophic (SQG) model seems to be appealing for the motions in both ocean and atmosphere. We shall see that the governing equations of SQG system consist of a transport equation with drift of the form (1.3), with the velocity $\boldsymbol{u}$ related to the solution of another partial differential equation involving spectral fractional Laplacian.

In this chapter we first provide a derivation of the governing equations to the surface quasi-geostrophic dynamics in the atmosphere in Section 5.1. The reduced model on the square Ω follows from a Stinga-Torrea representation consists of fractional Laplacian in Ω, as shown in Section 5.2. Numerical algorithms proposed in Section 5.3 are adapted from the approximation schemes developed in Chapters 3 and 4. Their numerical validations are provided in Section 5.4. Simulation results and discussions are provided in Section 5.5.

5.1 Mathematical model

We refer to [33, Chapter 6], [49] and [13] for detailed derivation of the governing equations to quasi-geostrophic system. In particular, the rescaling of variables for nondimensionalization are

addressed in Section 6.2 of [33] and Section 2 of [50].

The quasi-geostrophic system is put forward for small Rossby number $\varepsilon := U/(f_0 L)$, where U (units: $\mathrm{m\,s^{-1}}$) and L (units: m) are the horizontal scales for velocity and length respectively, and $f_0 = 2\omega \sin\phi_0$ (units: $\mathrm{m\,s^{-2}}$) is the Coriolis parameter at latitude ϕ_0 and ω the rate of rotation of the earth. The small Rossby number assumption requires ϕ_0 to be large, namely, we require the motion occurs in a mid-to-high latitude region, distant from the equator. Additionally, the horizontal velocities u and v are required to be independent of z, and the vertical velocity w vanishes for $z \to \infty$. Let $\boldsymbol{v} := (u, v, w)$ be the three-dimensional velocity vector field, with $u(x, y, t)$, $v(x, y, t)$, $w(x, y, z, t) : \mathcal{C} \times \mathbb{R}_+ \to \mathbb{R}$. We assume the fluid is incompressible and is of constant density, namely,

$$\nabla \cdot \boldsymbol{v} := \frac{\partial}{\partial x} u + \frac{\partial}{\partial y} v + \frac{\partial}{\partial z} w = 0, \qquad \text{for } (x, y, z) \in \mathcal{C},\ t \in \mathbb{R}_+. \tag{5.1}$$

With these assumptions the horizontal momentum equations are (see also [33, Formula 3.12.19])

$$\frac{\partial}{\partial t} u + u \frac{\partial}{\partial x} u + v \frac{\partial}{\partial y} u - f_0 v = -\frac{\partial}{\partial x} \Psi, \tag{5.2}$$

$$\frac{\partial}{\partial t} v + u \frac{\partial}{\partial x} v + v \frac{\partial}{\partial y} v + f_0 u = -\frac{\partial}{\partial y} \Psi, \tag{5.3}$$

both for $(x, y, z) \in \mathcal{C},\ t \in \mathbb{R}_+$, where $\Psi(x, y, z, t) : \mathcal{C} \times \mathbb{R}_+ \to \mathbb{R}$ denotes the geopotential. The hydrostatic relation between the potential temperature $\theta(x, y, z, t) : \mathcal{C} \times \mathbb{R}_+ \to \mathbb{R}$ and the geopotential Ψ is given by

$$\frac{\partial}{\partial z} \Psi = \frac{g\theta}{\theta_r}, \qquad \text{in} \quad \mathcal{C} \times \mathbb{R}_+. \tag{5.4}$$

Here g is the gravitational acceleration constant, θ_r is a reference potential temperature chosen to be $\theta(x, y, 0, t_0)$ at $z = 0$ for some $t_0 > 0$. Following from the first law of thermodynamics in the perfect gas with the absence of internal heating source, the conservation of potential temperature θ reads

$$\frac{\partial}{\partial t} \theta + \boldsymbol{v} \cdot \nabla \theta = 0, \qquad \text{in} \quad \mathcal{C} \times \mathbb{R}_+. \tag{5.5}$$

In summary, the complete dimensional system consists of (5.2), (5.3), (5.1), (5.4), and (5.5).

Non-dimensionalization

In order to fully understand the relationship between the parameters, we non-dimensionalize the governing equations by introducing non-dimensional variables denoted by $\hat{\cdot}$

$$x := L\hat{x}, \quad y := L\hat{y}, \quad z := D\hat{z}, \quad \hat{\Omega} := \{(\hat{x}, \hat{y}) : (x, y) \in \Omega\} = (0, 2\pi)^2,$$
$$t := \frac{L}{U}\hat{t}, \quad u := U\hat{u}, \quad v := U\hat{v}, \quad w := \frac{DU}{L}\hat{w}, \quad \Psi := f_0 U L \hat{\Psi},$$

where L and U are the horizontal length scale and velocity scale respectively appearing in the Rossby number, and D (units: m) is the vertical length scale, which is assumed to satisfy $D \ll L$. It in turn implies that the scale of vertical velocity is of order $O(DU/L) \ll 1$. This small relative vertical velocity assumption indicates that the potential temperature θ will be only slightly disturbed from the value it would be in the absence of motion. Therefore, θ admits a representation as the sum of the rest-state value $\theta_s(z)$ and a small time-dependent deviation of order $O(\varepsilon F)$ (c.f. [33, Formula 6.5.3])

$$\theta = \theta_s(z)\left(1 + \varepsilon F \hat{\theta}(x, y, z, t)\right),$$

with θ_s the "standard" temperature potential satisfying $\theta_s(0) = \theta_r$, and

$$F = \frac{f_0^2 L^2}{gD} = \left(\frac{L}{R}\right)^2, \qquad \text{with } R = \frac{\sqrt{gD}}{f_0}$$

denoting the external Rossby deformation radius; see [33, Formula 3.12.9]. In practice, large-scale waves in the atmosphere in mid-latitudes are characterized by

$$U = O(10\ m/s), \quad L = O(10^3\ km), \quad f_0 = O(10^{-4}\ s^{-1}), \quad D = O(10\ km).$$

This implies that $F = O(10^{-1}) = O(\varepsilon)$.

To simplify our notations, unless explicitly stated, from now on we drop the hat notation $\hat{\cdot}$ and

only consider the non-dimensional variables. With these notations and conventions, the continuity condition (5.1) remains

$$\frac{\partial}{\partial x}u + \frac{\partial}{\partial y}v + \frac{\partial}{\partial z}w = 0, \qquad \text{in} \quad \mathcal{C} \times \mathbb{R}_+, \tag{5.6}$$

the horizontal momentum equations (5.2) and (5.3) becomes

$$\varepsilon\left(\frac{\partial}{\partial t}u + u\frac{\partial}{\partial x}u + v\frac{\partial}{\partial y}u\right) - v = -\frac{\partial}{\partial x}\Psi, \tag{5.7}$$

$$\varepsilon\left(\frac{\partial}{\partial t}v + u\frac{\partial}{\partial x}v + v\frac{\partial}{\partial y}v\right) + u = -\frac{\partial}{\partial y}\Psi, \tag{5.8}$$

both in $\mathcal{C} \times \mathbb{R}_+$, the potential temperature and geopotential relation (5.4) on the lower boundary is rewritten as

$$\frac{\partial}{\partial z}\Psi = \frac{gD}{f_0UL}(1 + \varepsilon F\theta) = \frac{gD}{f_0UL} + \theta, \qquad \text{on} \quad \Omega \times \mathbb{R}_+, \tag{5.9}$$

and the potential temperature conservation condition (5.5) reads

$$\frac{\partial}{\partial t}\theta + u\frac{\partial}{\partial x}\theta + v\frac{\partial}{\partial y}\theta + w\frac{\partial}{\partial z}\theta = \frac{w}{\varepsilon F\theta_s}\frac{\partial\theta_s}{\partial z}(1 + \varepsilon F\theta), \qquad \text{in} \quad \mathcal{C} \times \mathbb{R}_+. \tag{5.10}$$

Equation simplification using $\varepsilon-$expansions

For small $\varepsilon \ll 1$, we expand the velocity $u(x, y, \varepsilon, t)$ as (recall that u is independent of z)

$$u(x, y, \varepsilon, t) = u_0(x, y, t) + \varepsilon u_1(x, y, t) + o(\varepsilon)u_2(x, y, t), \tag{5.11}$$

with u_0, u_1, $u_2 : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ functions independent of ε and of magnitude $O(1)$. The other functions v, w, and Ψ are expanded in the same way. Terms in the equations of the same order in ε should be balanced. We now discuss the corresponding $O(1)$ and $O(\varepsilon)$ approximations of equations (5.6) to (5.10).

We start with (5.7) and (5.8). The $O(1)$ order approximations to (5.7) and (5.8) yield the

geostrophic relations

$$u_0 = -\frac{\partial}{\partial y}\Psi_0, \qquad v_0 = \frac{\partial}{\partial x}\Psi_0, \qquad \text{both in} \quad \Omega \times \mathbb{R}_+, \tag{5.12}$$

which immediately imply that

$$\frac{\partial}{\partial x}u_0 + \frac{\partial}{\partial y}v_0 = 0.$$

This, together with (5.6) yields the $O(\varepsilon)$ approximation

$$\varepsilon\left(\frac{\partial}{\partial x}u_1 + \frac{\partial}{\partial y}u_1 + \frac{\partial}{\partial z}w_1\right) + \frac{\partial}{\partial z}w_0 = 0, \qquad \text{in} \quad \mathcal{C} \times \mathbb{R}_+.$$

It follows that

$$\frac{\partial}{\partial x}u_1 + \frac{\partial}{\partial y}u_1 + \frac{\partial}{\partial z}w_1 = 0. \tag{5.13}$$

and $\partial_z w_0 = 0$. Furthermore, because it is assumed that $\lim_{z\to\infty} w = 0$, it enforces $w_0 = 0$.

The $O(\varepsilon)$ terms in (5.7) and (5.8) imply:

$$\frac{\partial}{\partial t}u_0 + u_0\frac{\partial}{\partial x}u_0 + v_0\frac{\partial}{\partial y}u_0 - v_1 = -\frac{\partial}{\partial x}\Psi_1, \tag{5.14}$$

$$\frac{\partial}{\partial t}v_0 + u_0\frac{\partial}{\partial x}v_0 + v_0\frac{\partial}{\partial y}v_0 + u_1 = -\frac{\partial}{\partial y}\Psi_1, \tag{5.15}$$

both in $\Omega \times \mathbb{R}_+$. In order to eliminate the potential Ψ_1, we differentiate (5.14) with respect to y, and (5.15) with respect to x, and subtract the resulting equations to obtain the vorticity equation

$$\frac{\partial}{\partial t}\zeta_0 + u_0\frac{\partial}{\partial x}\zeta_0 + v_0\frac{\partial}{\partial y}\zeta_0 = -\left(\frac{\partial}{\partial x}u_1 + \frac{\partial}{\partial y}v_1\right) = \frac{\partial}{\partial z}w_1, \qquad \text{in} \quad \Omega \times \mathbb{R}_+. \tag{5.16}$$

where we applied (5.13) for the last equality. Here ζ_0 is the relative vorticity, and is given by

$$\zeta_0 := \frac{\partial}{\partial x}v_0 - \frac{\partial}{\partial y}u_0 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\Psi_0, \tag{5.17}$$

in view of (5.12).

We now focus on (5.10) (recall $w_0 = 0$), the $O(1)$ terms imply

$$\frac{\partial}{\partial t}\theta_0 + u_0\frac{\partial}{\partial x}\theta_0 + v_0\frac{\partial}{\partial y}\theta_0 = -S(z)w_1, \qquad \text{in} \quad \Omega \times \mathbb{R}_+, \tag{5.18}$$

where

$$S(z) = \frac{1}{F\theta_s}\frac{\partial \theta_s}{\partial z} = \frac{N_s^2(z)D^2}{f_0^2 L^2}, \qquad \text{with } N_s^2(z) := \frac{1}{D}\frac{g}{\theta_s}\frac{\partial \theta_s}{\partial z}.$$

Here $N_s(z)$ is called the Brunt-Väisälä frequency of the rest-state atmosphere; see [51, 52, 33]. A typical magnitude of N_s in the atmosphere is $O(10^{-2}\ s^{-1})$. Therefore, the the stratification parameter $S(z)$ is of order $O(1)$. The boundary-layer theory (c.f. Section 4.5 of [33]) indicates that the vertical velocity pumped out of the lower boundary Ekman layer satisfies (see [33, Formulas 4.5.39 and 6.6.8])

$$w(x,y,0) = \varepsilon w_1(x,y,0) = \frac{E_V^{1/2}}{2}\left(\frac{\partial}{\partial x}v_0 - \frac{\partial}{\partial y}u_0\right) = \frac{E_V^{1/2}}{2}\zeta_0 = \frac{E_V^{1/2}}{2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\Psi_0,$$

where $E_V = 2A_V/(f_0 D^2)$ is the vertical Ekman number, and A_V is the vertical turbulent viscosity coefficient (c.f. [33, Formula 6.2.31]). Our main point of interest relies on the small Ekman pumping, therefore we require $E_V^{1/2}/(2\varepsilon) \ll 1$. The above identity combined with (5.18), yield

$$\frac{\partial}{\partial t}\theta_0 + \boldsymbol{u}\cdot\nabla\theta_0 = S(z)\frac{E_V^{1/2}}{2\varepsilon}(-\Delta_\Omega)\Psi_0, \qquad \text{in} \quad \Omega \times \mathbb{R}_+, \tag{5.19}$$

where

$$\boldsymbol{u} := (u_0, v_0) = \left(-\frac{\partial}{\partial y}\Psi_0, \frac{\partial}{\partial x}\Psi_0\right). \tag{5.20}$$

The $O(1)$ approximations of $\Psi = \Psi_0 + O(\varepsilon)\Psi_1$ and $\theta = \theta_0 + O(\varepsilon)\theta_1$ in (5.9) yield

$$\frac{\partial}{\partial z}\Psi_0 = \frac{gD}{f_0 UL} + \theta_0, \qquad \text{in} \quad \mathcal{C} \times \mathbb{R}_+. \tag{5.21}$$

The relation can be further simplified by absorbing the constant term in the right hand side through

redefining $\theta_0 := gD/(f_0UL) + \theta_0$. For this reason we write

$$\frac{\partial}{\partial z}\Psi_0 = \theta_0, \qquad \text{on} \quad \Omega \times \mathbb{R}_+. \tag{5.22}$$

We introduce the potential vorticity $\mathcal{G} := \zeta_0 + \frac{\partial}{\partial z}(\frac{\theta_0}{S(z)})$ to eliminate the vertical velocity w_1 in (5.16) and (5.18) and obtain (c.f. Equation 6.5.21 in [33])

$$\frac{\partial}{\partial t}\mathcal{G} + \boldsymbol{u} \cdot \nabla \mathcal{G} = 0, \qquad \text{in} \quad \Omega \times \mathbb{R}_+. \tag{5.23}$$

Such conservation of potential vorticity serves as a critical property of the quasi-geostrophic system (see, e.g., [53, 33, 54]). In view of (5.21) and (5.17), we realize that $\mathcal{G}$ satisfies

$$\mathcal{G} = \frac{\partial^2}{\partial x^2}\Psi_0 + \frac{\partial^2}{\partial y^2}\Psi_0 + \frac{\partial}{\partial z}\Big(\frac{1}{S(z)}\frac{\partial}{\partial z}\Psi_0\Big). \tag{5.24}$$

In particular, if we assume a constant Brunt-Väisälä frequency $N_s(z) = N$ (i.e., $S(z) = S$ is assumed to be independent of z), we can absorb the factor $1/S$ through rescaling the vertical variable z with $\tilde{z} := S^{1/2}z$. In what follows, we retain the notation z for the rescaled vertical coordinate, and write

$$\mathcal{G} = \Delta\Psi_0. \tag{5.25}$$

The special case $\mathcal{G} = 0$, often referred as surface quasi-geostrophic (SQG) flow, corresponds to uniform potential vorticity in the interior of the fluid; cf. [13, Section 2] and [55]. This simplification allows to determine the flow entirely by the surface potential temperature $\theta_\Omega := \theta(x, y, 0, t)$. To sum up, when $N_s(z)$ is constant, and $\mathcal{G} \equiv 0$, we obtain the non-dimensional equations

$$\frac{\partial}{\partial t}\theta_\Omega + \boldsymbol{u} \cdot \nabla\theta_\Omega - \kappa(-\Delta_\Omega)\Psi_\Omega = 0, \quad \text{on} \quad \Omega \times \mathbb{R}_+, \tag{5.26}$$

$$\Delta\Psi_0 = 0, \quad \text{in} \quad \Omega \times (0, \infty) \times \mathbb{R}_+, \tag{5.27}$$

$$\partial_z\Psi_0 = \theta_\Omega, \quad \text{on} \quad \Omega \times \mathbb{R}_+, \tag{5.28}$$

where $\Psi_\Omega := \Psi_{0|\Omega}$, and the dissipation coefficient $\kappa = SE_V^{1/2}/(2\varepsilon) \ll 1$ is independent of x, y, and t. Here thanks to (5.20), the two-dimensional velocity $\boldsymbol{u}$ is determined by

$$\boldsymbol{u} = \Big(-\frac{\partial}{\partial y}\Psi_0|_{z=0}, \frac{\partial}{\partial x}\Psi_0|_{z=0}\Big). \tag{5.29}$$

We additionally impose the decay condition

$$\lim_{z\to\infty} \Psi_0(x,y,z) = 0, \quad \forall (x,y) \in \Omega \tag{5.30}$$

to close the system.

The temperature potential θ_Ω and the potential Ψ_Ω are supplemented with periodic boundary condition, i.e., $\theta_\Omega,\ \Psi_\Omega \in H_p^{2s}(\Omega)$, where

$$H_p^{2s}(\Omega) := \left\{ u \in H^{2s}(\Omega)\ :\ u(x,0) = u(x,2\pi), \text{ and } u(0,y) = u(2\pi,y) \right\}. \tag{5.31}$$

To close the system we impose the mean value conditions

$$\int_\Omega \theta_\Omega(t)\, dx = \int_\Omega \theta_\Omega(0)\, dx, \quad \text{and } \int_\Omega \Psi_\Omega(t)\, dx = 0, \quad t > 0.$$

5.2 Model reduction

Notice that although in (5.29) we only require the trace of Ψ on the 2D domain Ω, the potential function Ψ_0 has to be solved in the 3D space. We now replace (5.27), (5.28), and (5.30) by a nonlocal problem on the two-dimensional domain Ω.

It turns out that Ψ_Ω solves

$$(-\Delta_\Omega)^{1/2}\Psi_\Omega = -\theta_\Omega \quad \text{in } \Omega. \tag{5.32}$$

Here the operator $(-\Delta_\Omega)^{1/2}$ is spectral fractional laplacian defined by (1.1) but associated with

periodic boundary condition. The proof is based on the discrete spectrum expansion. The concrete derivation is essentially the same as in Section 3 of [21] but replacing the eigenpairs to $-\Delta$ with homogeneous Dirichlet boundary condition on $\partial\Omega$ by the eigenpairs associated to $-\Delta_\Omega$ with periodic boundary condition. The proof is thus omitted in this work.

From now on we drop the subscript Ω for notational convenience. Substitute (5.32) back into (5.26), we obtain the restricted model

$$\frac{\partial}{\partial t}\theta + \boldsymbol{u}\cdot\nabla\theta + \kappa(-\Delta)^{1/2}\theta = 0, \quad \boldsymbol{u} = \nabla^\perp\Psi_\Omega, \quad (-\Delta)^{1/2}\Psi_\Omega = -\theta \quad \text{all in } \Omega. \tag{5.33}$$

In [55, 1] the Ekman pumping term in the above SQG equation is often generalized by considering fractional powers $s \in [1/2, 1)$. In the remaining of this chapter we will consider the following generalized SQG equation:

$$\frac{\partial}{\partial t}\theta + \boldsymbol{u}\cdot\nabla\theta + \kappa(-\Delta)^{s}\theta = 0, \quad \boldsymbol{u} = \nabla^\perp\Psi_\Omega, \quad (-\Delta)^{1/2}\Psi_\Omega = -\theta \quad \text{all in } \Omega, \tag{5.34}$$

with the fractional power $s \in [1/2, 1)$, and $0 \le \kappa \ll 1$.

For the inviscid case $\kappa = 0$, the conservation laws indicate that the kinetic energy $E_k(\theta)$ and helicity $H(\theta)$ are conserved throughout time $t \in [0, \infty]$:

$$E_k(\theta) = \frac{1}{2}\int_\Omega \theta(x,y,t)^2 dx, \qquad H(\theta) = -\int_\Omega \Psi_\Omega(x,y,t)\theta(x,y,t)\, dx. \tag{5.35}$$

Indeed, to derive the conservation of kinetic energy, we multiply both sides of (5.34) by θ and integrate on Ω. From (3.7) we have $\int_\Omega (\boldsymbol{u}\cdot\nabla\theta)\theta = 0$, and thus

$$0 = \int_\Omega \frac{\partial}{\partial t}\theta(x,y,t)\theta(x,y,t) = \frac{1}{2}\frac{\partial}{\partial t}\int_\Omega \theta(x,y,t)^2.$$

The second conservation property is derived upon observing

$$H(\theta) = \int_\Omega \big((-\Delta)^{-1/2}\theta\big)\, \theta \, dx.$$

5.3 Numerical approximation scheme

In this section we put forward our numerical algorithm to approximate (5.33). Let $\tau > 0$ be the time stepping parameter. Given an initial condition for the potential temperature $\theta(x, y, 0) = \theta_0(x, y)$, we look for approximations to $\theta(t_n)$ recursively for $t_n = n\tau$, $n = 1, 2, \cdots$. At each time step t_{n+1}, suppose $\theta^n : \Omega \to \mathbb{R}$ is given, the solution $\theta(t_{n+1})$ to (5.33) is approximated with $\theta^{n+1} : \Omega \to \mathbb{R}$ by invoking the operator splitting (c.f. [56]) together with the forward Euler time discretization. This requires solving the three equations described below.

$$\theta^{n+1/2} - \theta^n + \tau(-\Delta)^s\theta^n = 0, \tag{5.36}$$

$$\Psi^{n+1/2} = (-\Delta)^{-1/2}\theta^{n+1/2}, \qquad \boldsymbol{u}^{n+1/2} = \Big(-\frac{\partial}{\partial y}\Psi^{n+1/2}, \frac{\partial}{\partial x}\Psi^{n+1/2}\Big), \tag{5.37}$$

$$\theta^{n+1} - \theta^{n+1/2} + \tau \boldsymbol{u}^{n+1/2}\nabla\theta^{n+1/2} = 0. \tag{5.38}$$

A finite element method is further advocated for the space discretization. We denote $\{\mathcal{T}_h\}_{h>0}$ to be a sequence of quasi-uniform and shape regular polygonal partition of Ω. For a fixed $h > 0$, for any given $T \in \mathcal{T}_h$, we denote

$$F_T : [0, 1]^2 \to T \tag{5.39}$$

to be the bilinear mapping that maps the reference element (the unit square) to the actual element T. With this mapping we define the finite element space

$$\begin{aligned} V_h := \{v \in C^0(\overline{\Omega}) \ &: v_{|T} \circ F_T \text{ is bilinear for all } T \in \mathcal{T}_h, \\ &v(x, 0) = v(x, 2\pi),\ v(0, y) = v(2\pi, y), \text{ and } \int_\Omega v = 0\}. \end{aligned}$$

We now describe the fully discrete schemes to the three equations (5.36) to (5.38) separately.

Step 1: the intermediate temperature potential $\theta_h^{n+1/2}$

In this step we approximate for (5.36). We set the initial condition $\theta_h^0 := \pi_h\theta^0$, with π_h the $L^2(\Omega)$ projection onto V_h. Recursively, assume that $\theta_h^n \in V_h(\Omega)$ approximating θ^n is obtained, we seek $\theta_h^{n+1/2} \in V_h(\Omega)$.

Since the fractional term in (5.36) is explicit, it requires the approximation of $\int_\Omega(-\Delta)^s\theta_h^n\varphi_h$, for $\varphi_h \in V_h$. The approach is essentially the same as in Chapter 3, which consists of a sinc scheme to the integral representation, followed by a finite element discretization. The resulting computable approximation reads

$$\int_\Omega(-\Delta)^s\theta_h^n\varphi_h \approx a_p^h(\theta_h^n,\varphi_h) := \frac{\sin(\pi s)}{\pi}k\sum_{j=-N_-}^{N_+} e^{-sy_j}\int_\Omega(\eta_h(y_j)+\theta_h^n)\varphi_h, \tag{5.40}$$

where k is the sinc quadrature stepping, $y_j = jk$, $N_+ := \left\lceil\frac{2\pi^2}{sk^2}\right\rceil$, and $N_- := \left\lceil\frac{4\pi^2}{(1-s)k^2}\right\rceil$. Here for each y_j, $\eta_h(y_j) := \eta_h(y_j;\theta_h^n) \in V_h$ solves

$$\int_\Omega \eta_h(y_j)v_h + e^{y_j}\int_\Omega \nabla\eta_h(y_j)\nabla v_h = -\int_\Omega \theta_h^n v_h, \qquad \text{for all } v_h \in V_h.$$

Returning to the approximation of (5.36), we apply a Strongly Stability Preserving two-stage Runge-Kutta (SSP-RK2) scheme [35] to ensure a second order convergence in time. In the first stage we seek solution $\mu_h^{(1)} \in V_h$ such that

$$\int_\Omega \mu_h^{(1)}\varphi_h = -\tau\kappa a_p^h(\theta_h^n,\varphi_h) + \int_\Omega \theta_h^n\varphi_h, \qquad \text{for all } \varphi_h \in V_h. \tag{5.41}$$

In the second stage, we solve for $\mu_h^{(2)} \in V_h$ satisfying (5.41) but with θ_h^n replaced by $\mu_h^{(1)}$ in the right hand side. We then define the intermediate approximation to the temperature potential by

$$\theta_h^{n+1/2} := \frac{1}{2}\left(\theta_h^n + \mu_h^{(2)}\right). \tag{5.42}$$

Step 2: the new velocity $\boldsymbol{u}_h^{n+1/2}$

In this step we update the geopotential, from which we update the velocity. For $\theta_h^{n+1/2} \in V_h$ obtained in Step 1, the finite element approximation to $\Psi_{h,\Omega}^{n+1/2} := (-\Delta)^{-1/2}\theta_h^{n+1/2}$ could be treated the same as in Section 2.7, namely,

$$\Psi_{h,\Omega}^{n+1/2} \approx \Psi_h^{n+1/2} := \frac{2k}{\pi} \sum_{j=-N_-}^{N_+} e^{z_j} \eta_h(z_j), \tag{5.43}$$

where k is the sinc stepping, $z_j = jk$, $N_+ := \left\lceil \frac{\pi^2}{k^2} \right\rceil$, and $N_- := \left\lceil \frac{\pi^2}{k^2} \right\rceil$. The function $\eta_h(z_j) := \eta_h(z_j; \theta_h^{n+1/2}) \in V_h$ satisfies

$$\int_\Omega \eta_h(z_j) v_h + e^{2z_j} \int_\Omega \nabla \eta_h(z_j) \nabla v_h = \int_\Omega \theta_h^{n+1/2} v_h, \qquad \text{for all } v_h \in V_h.$$

Hence, the velocity $\boldsymbol{u}^{n+1/2}$ is approximated as

$$\boldsymbol{u}^{n+1/2} \approx \boldsymbol{u}_h^{n+1/2} := \left(-\frac{\partial}{\partial y} \Psi_h^{n+1/2}, \frac{\partial}{\partial x} \Psi_h^{n+1/2} \right).$$

Step 3: the new temperature potential θ_h^{n+1}

The last step approximates (5.38). The same SSP-RK2 method as in Subsection 5.3 is applied to ensure a second order convergence in time. For the given $\theta_h^{n+1/2}$ and $\boldsymbol{u}_h^{n+1/2}$ from Steps 1 and 2 respectively, in the first stage we seek $\mu_h^{(1)} \in V_h$ such that

$$\int_\Omega \mu_h^{(1)} \varphi_h = \int_\Omega \theta_h^{n+1/2} \varphi_h + \tau \int_\Omega (\boldsymbol{u}_h^{n+1/2} \cdot \nabla \theta_h^{n+1/2}) \varphi_h, \qquad \text{for all } \varphi_h \in V_h. \tag{5.44}$$

In the following second stage, we solve for $\mu_h^{(2)} \in V_h$ solving (5.44) but with $\theta_h^{n+1/2}$ replaced by $\mu_h^{(1)}$ in the right hand side. The final approximation $\theta_h^{n+1} \in V_h$ to θ^{n+1} reads

$$\theta_h^{n+1} := \frac{1}{2} \left(\theta_h^{n+1/2} + \mu_h^{(2)} \right). \tag{5.45}$$

It is well documented that the finite element approximations of transport equation might be polluted by spurious oscillations. To circumvent this issue, we include the smoothness-based second-order maximum principle preserving viscosity method proposed in [36]. The numerical parameters used in the artificial viscosity are those recommended in Formula (5.4) in [57].

5.4 Numerical scheme validation

In all cases the domain $\Omega = (0, 2\pi) \times (0, 2\pi) \subset \mathbb{R}^2$.

The parabolic fractional Laplacian scheme in Subsection 5.3 is validated using for $m,\ n \in \mathbb{N}$,

$$\theta(x,t) = e^{-\kappa(m^2+n^2)^s t}\cos(nx)\sin(my),$$

which implies $\partial_t\theta + (-\Delta)^s\theta = 0$, and $\theta(x, y, 0) = \cos(nx)\sin(my)$. Notice that we use the fact that $\cos(nx)\sin(my)$ is an eigenfunction of the periodic Laplacian on $(0, 1)^2$ with eigenvalue $m^2 + n^2$. We take $m = 1$, $n = 2$, $s = 0.4$, $\kappa = 0.001$, $\tau = 0.002$, a sinc quadrature size $k = 0.2$, and a final time $T = 20$. Here τ and k are chose to be small enough such that the sinc quadrature error and time discretization error do not influence the space discretization error. The domain Ω is uniformly divided into small squares of sizes $h_i = 1/2^i$ for $i = 5,\ 6,\ 7,\ 8$.

The L^2 error $e_i := \left\|\theta(t_n) - \theta^n_{h_i}\right\|$ and the observed rate of convergence

$$ORC_i := \ln(e_i/e_{i+1})/\ln(h_i/h_{i+1})$$

are reported in Table 5.1. Due to the smoothing property of parabolic equation [30], the theoretical convergence rate for space discretization is h^2. This is indeed what is observed.

We next verify the accuracy of approximating transport equation using the scheme (5.3), where $\boldsymbol{u}_h^{n+1/2} = (U_1, U_2)$ is constant in time. This transport equation admits the exact solution

$$\theta(x, y; t) = \theta_0(x - U_1 t, y - U_2 t).$$

	$t=1$		$t=5$		$t=10$		$t=20$	
h	L^2 error	ORC	L^2 error	ORC	L^2 error	ORC	L^2 error	ORC
1/32	5.339E-02		5.308E-02		5.270E-02		5.194E-02	
1/64	1.341E-02	1.994	1.333E-02	1.994	1.323E-02	1.994	1.304E-02	1.993
1/128	3.356E-03	1.998	3.336E-03	1.998	3.312E-03	1.998	3.265E-03	1.998
1/256	8.392E-04	2.000	8.344E-04	2.000	8.284E-04	1.999	8.166E-04	1.999

Table 5.1. L^2 error and ORC for parabolic equation. The observed convergence rate matches the theoretical rate of convergence of 2.

In particular, the numerical method is implemented with

$$\theta_0(x) = \sin(x)\sin(y) + \cos(y),$$

and $U_1 = U_2 = 1$. The numerical parameters are the time step $\tau = 0.002$, the final time $T = 20$, and the uniformly divided sundivisions of sizes $h_i = 1/2^i$ for $i = 5,\ 6,\ 7,\ 8$. We summarize the results in Table 5.2 at various times. The theoretical convergence rate of order 2 for the L^2 error $e_i := \left\|\theta(t_n) - \theta^n_{h_i}\right\|$ is observed.

	$t=1$		$t=5$		$t=10$		$t=20$	
h	L^2 error	ORC	L^2 error	ORC	L^2 error	ORC	L^2 error	ORC
1/32	1.251E-01		4.790E-01		8.178E-01		1.320E+00	
1/64	3.003E-02	2.059	1.194E-01	2.004	2.180E-01	1.907	4.001E-01	1.722
1/128	7.009E-03	2.099	2.834E-02	2.074	5.281E-02	2.045	9.918E-02	2.012
1/256	1.622E-03	2.112	6.655E-03	2.091	1.257E-02	2.071	2.398E-02	2.048

Table 5.2. L^2 error and ORC for the transport equation. The observed convergence rate matches the theoretical rate of convergence of 2.

5.5 Numerical results

In this section the efficiency and accuracy of our approximation scheme is tested for the computation of solutions corresponding to the following two different smooth initial conditions

$$\theta_0(x) = \sin(x)\sin(y) + \cos(y), \tag{5.46}$$

$$\theta_0(x) = \exp\left(-(x-\pi)^2 - 16(y-\pi)^2\right). \tag{5.47}$$

The first initial condition, namely a linear combination of two eigenfunctions, leads to a front formation with a hyperbolic saddle, which exhibits a singular behavior (see also [1, 34, 55, 58]). The ability of capturing the rapidly growing high gradient region is crucial for the success of our numerical scheme. In order to record the time evolution, we plot the contour lines of the solution θ at each time step. Various literatures (c.f. [1, 34, 55, 58]) suggest that the behaviors at $t = 6,\ 8,\ 14,\ 20$ are of main interest. In addition, for fractional power $s < 1/2$ and dissipation parameter $\kappa > 0$, the question of whether the solution θ develops finite-time singularity remains open. We study numerically the finite-time behavior by setting $\kappa = 0.001$ and letting s vary within the range $[0.4, 0.5]$. The results are reported in Subsection 5.5.1.

The second initial condition, an ellipse of eccentricity 4, depicts a fast spinning flat vortex tending to shed filaments. It mimics the cyclonic circulations within the atmosphere in reality. The results are reported in Subsection 5.5.2.

5.5.1 Numerical results using (5.46) as initial condition

For the simulation with initial condition (5.46), we set grid size to be $h = 1/512$, time stepping $\tau = 0.002$, sinc quadrature stepping $k = 0.4$, and final time $T = 20$. We distinguish the inviscid ($\kappa = 0$) and viscid-limit ($\kappa = 0.001$) cases.

We report the contours at $t = 6,\ 8,\ 20$ for the case $\kappa = 0$ in Figure 5.1 in the left column. For comparison purpose, we also include the results from [1] in the right column of Figure 5.1. For all cases our contours are in good agreement.

For inviscid SQG with $\kappa = 0.001$, we choose fractional power $s = 0.4$ and report in Figure 5.2 the contour lines for the solution θ at $t = 6,\ 8,\ 20$ (left column) as well as the reference results obtained from [2] (right column); Our contours match the existing results very well. In addition, a comparison between Figure 5.1 and Figure 5.2 reveals that the time evolution in inviscid and viscid-limit cases are similar.

In our last set of experiments we fix $\kappa = 0.001$ and compute the solutions for various $s = 0.4,\ 0.45,\ 0.48,\ 0.5$. The corresponding contours at $t = 14$ are plotted in Figure 5.3. The results indicate that the value $s = 0.5$ does not appear to be critical for the behavior of solutions.

Throughout all simulations, we monitor the kinetic energy $E_k(\theta)$ and helicity $H(\theta)$ (defined in (5.35)). Due to the existence artificial viscosity, the decay of energies is inevitable. As illustrated by Figure 5.4, for the inviscid SQG at $t = 20$, the helicity and kinetic energy experience decay rates of 0.55% and 3.60% respectively.

5.5.2 Numerical results using (5.47) as initial condition

We only consider the inviscid case $\kappa = 0$. The mesh size is $h = 1/2048$, the time stepping τ is 0.002, the sinc quadrature stepping k is 0.4, and final time T is 40.

Figure 5.6 represents the numerical results at time $t = 8,\ 16,\ 26,\ 35$. These simulations are consistent with the results in [13, 2]. In particular, the smaller vortices close to the major structure are well-captured for $t \geq 26$. Throughout the filamentation process, the vortex retains its eccentricity. The pointing directions of the main axis of the ellipse are in good agreement with the results in [13, 2].

As illustrated in Figure 5.5, at $t = 40$ the helicity and kinetic energy decays by 0.80% and 5.05% respectively.

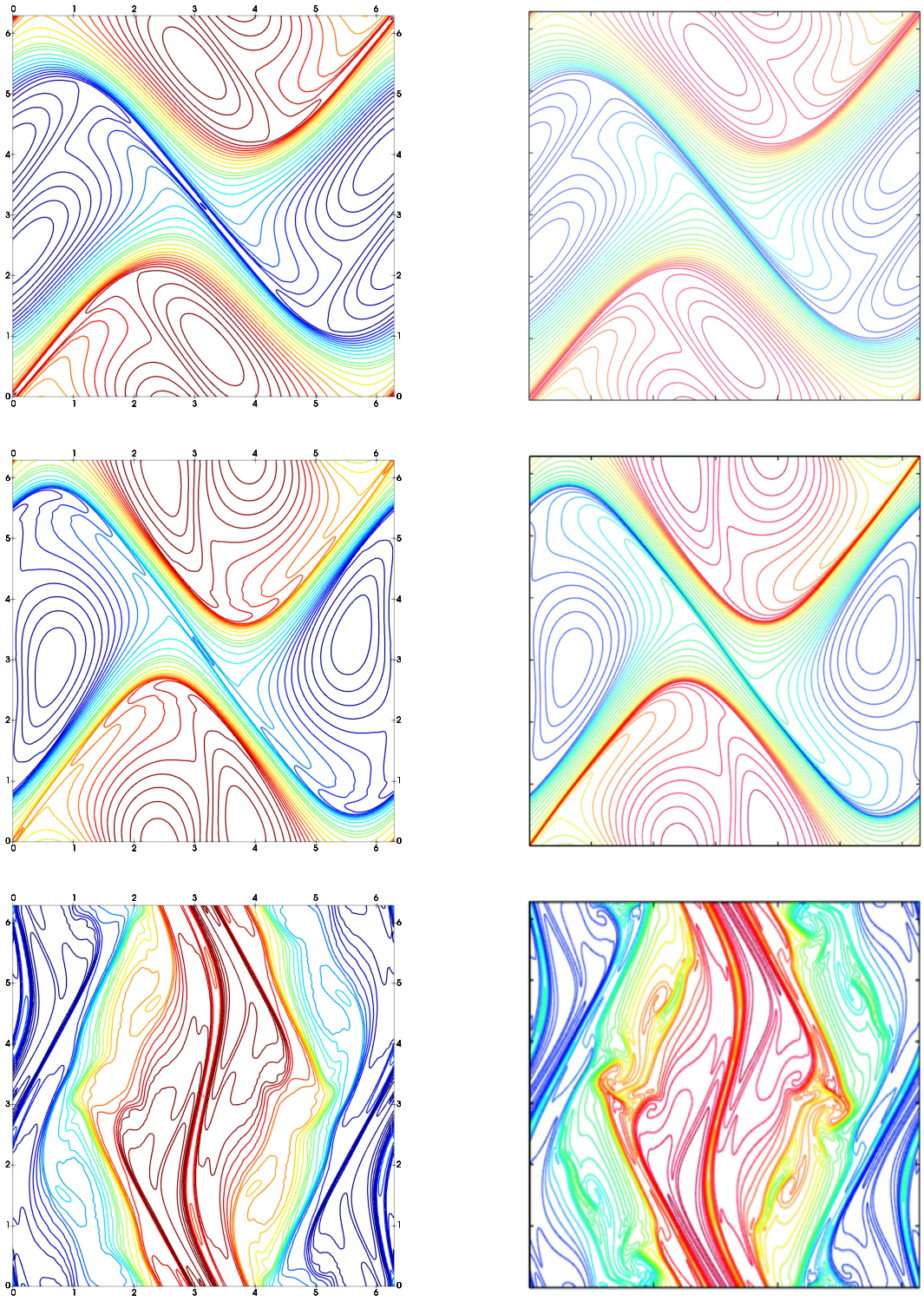


Figure 5.1. Contours of potential temperature θ for $\kappa = 0$ at $t = 6,\ 8,\ 20$ from top to bottom with initial condition defined by (5.46). Our simulation results (left column) are compared with results reprinted from Figures 3, 6, and 8 in [1] (right column).

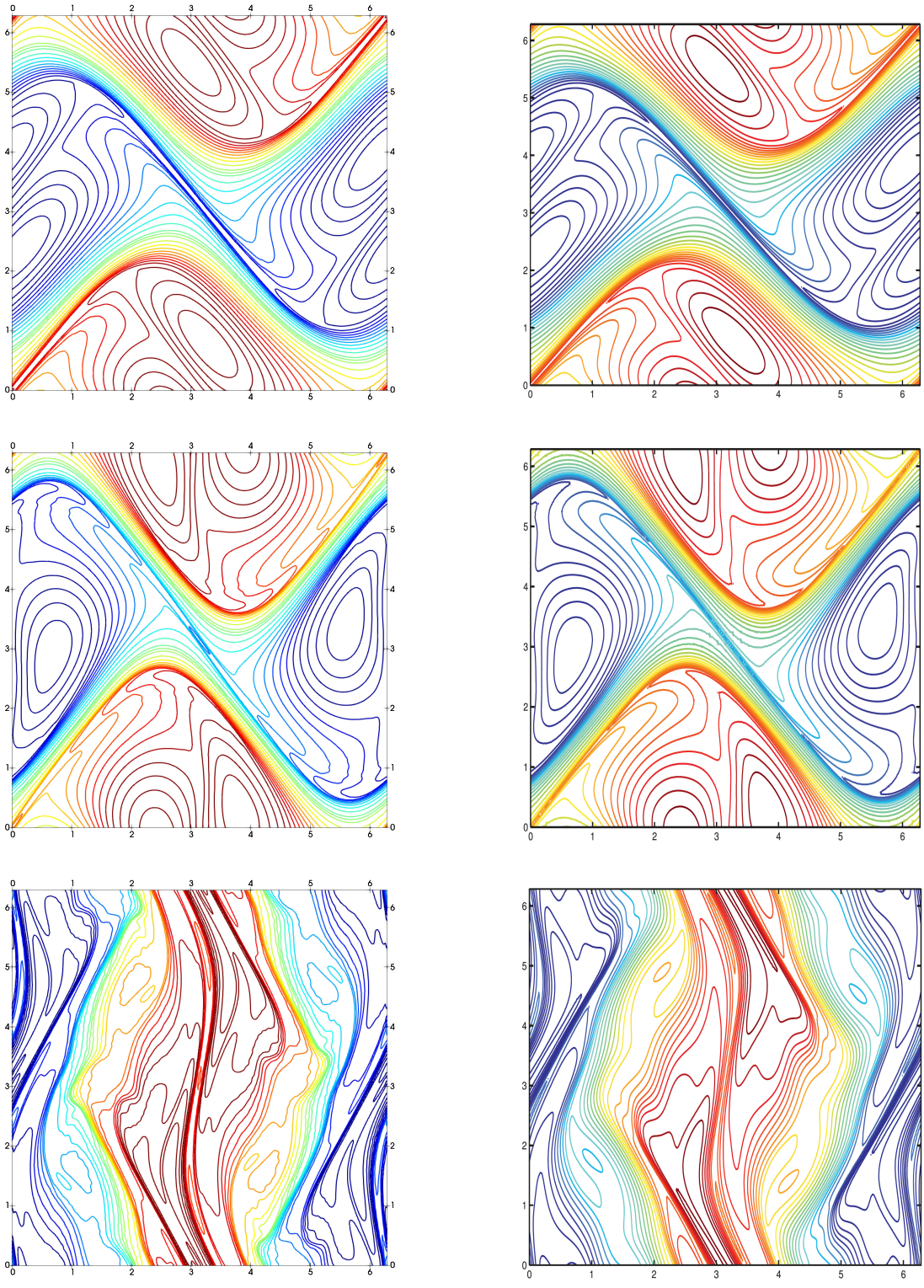
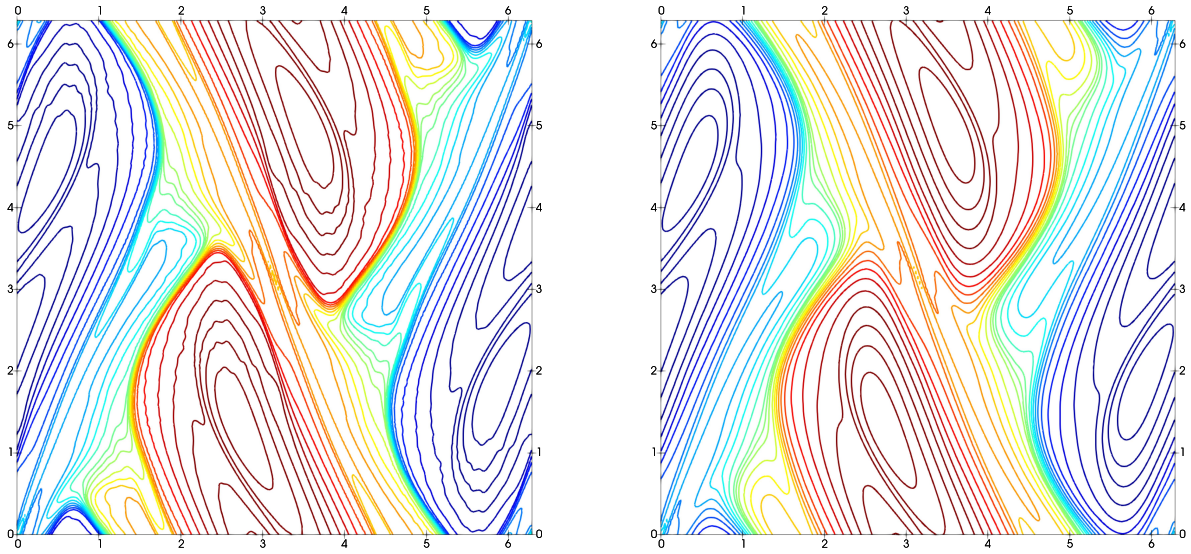
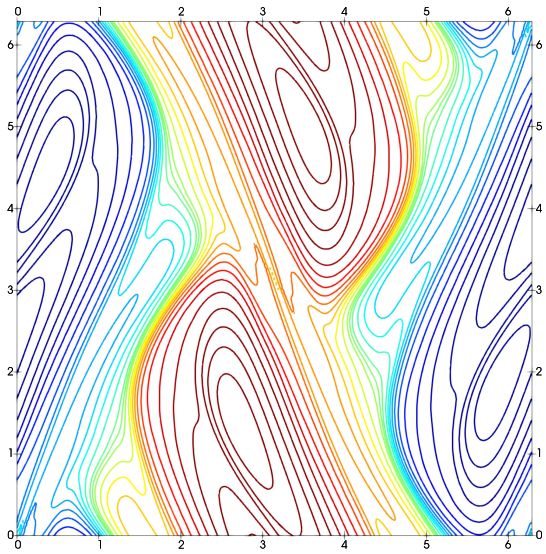


Figure 5.2. Contours of potential temperature θ for $\kappa = 0.001$, $s = 0.4$ at $t = 6,\ 8,\ 20$ from top to bottom with initial condition defined by (5.46). Our simulation results (left column) are compared with results reprinted from Figures 4, 5, and 7 in [2] (right column).

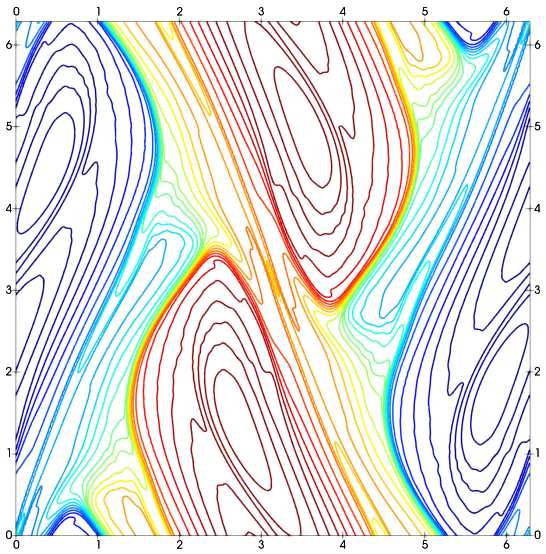


(a)

(b)

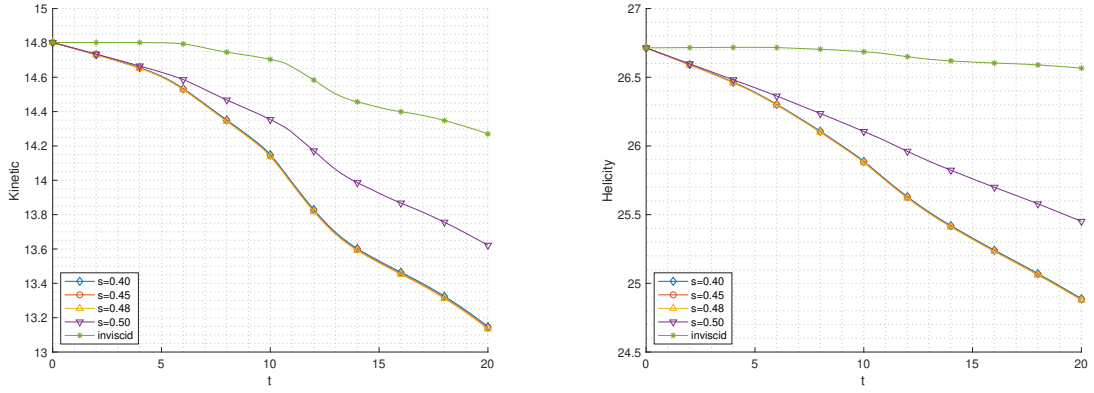


(c)



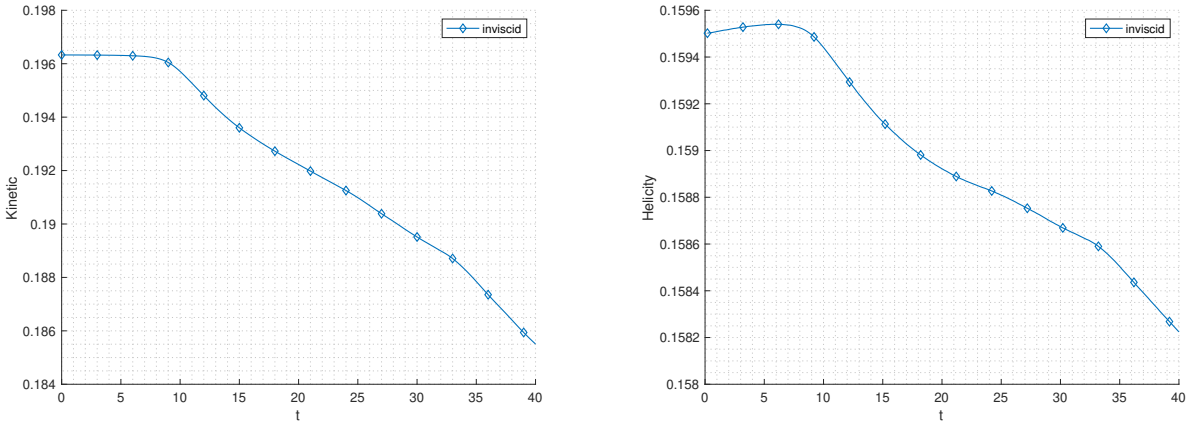
(d)

Figure 5.3. Comparison of contours for viscid-limit SQG ($\kappa = 0.001$) at $t = 14$ for various fractional power s: (a) $s = 0.4$, (b) $s = 0.45$, (c) $s = 0.48$, (d) $s = 0.5$. Initial condition is take as (5.46).



(a) (b)

Figure 5.4. The kinetic energy (a) and helicity (b) evolution with respect to time for various settings with initial condition defined by (5.46).



(a) (b)

Figure 5.5. The kinetic energy (a) and helicity (b) evolution respect to time for inviscid SQG with initial condition defined by (5.47).

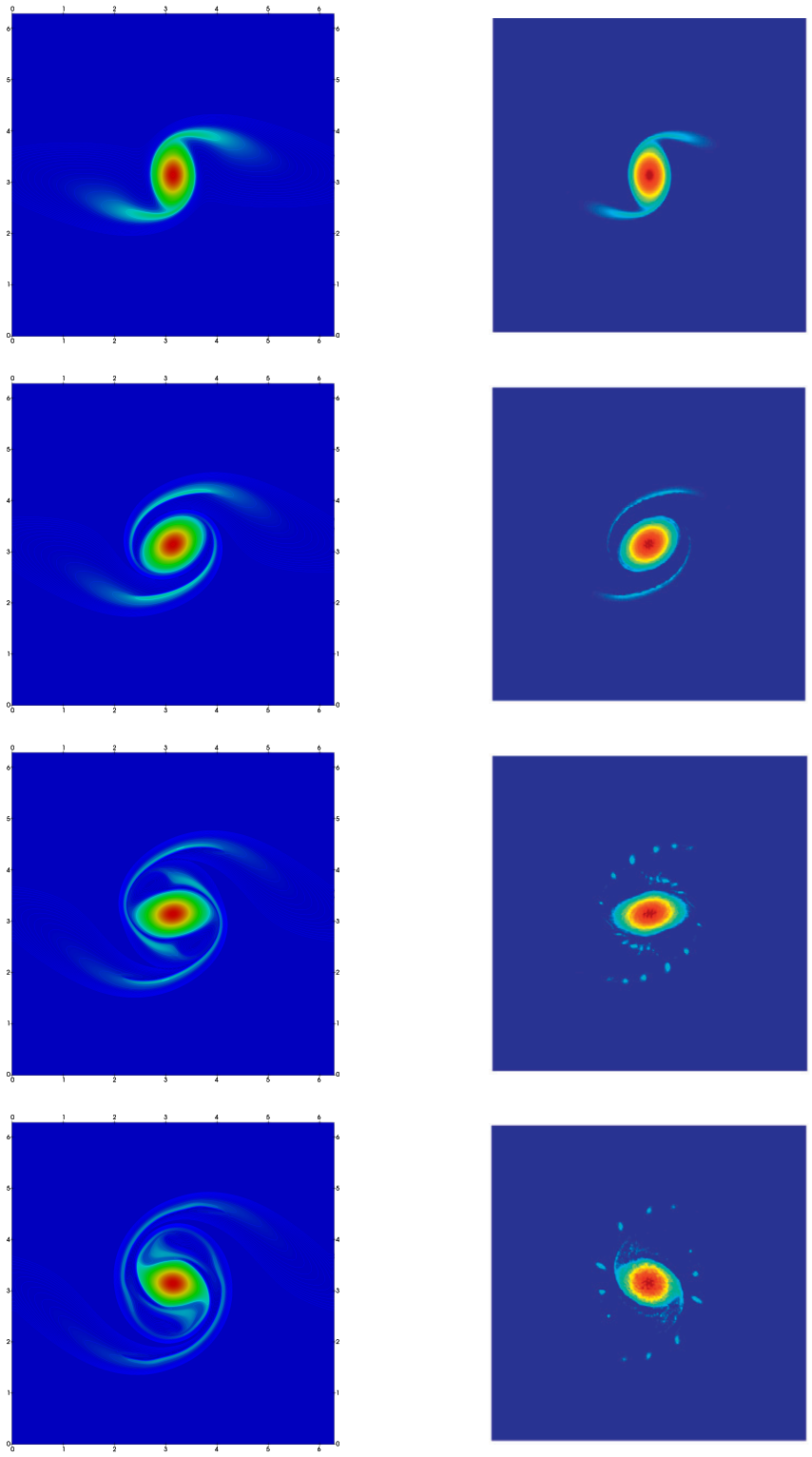


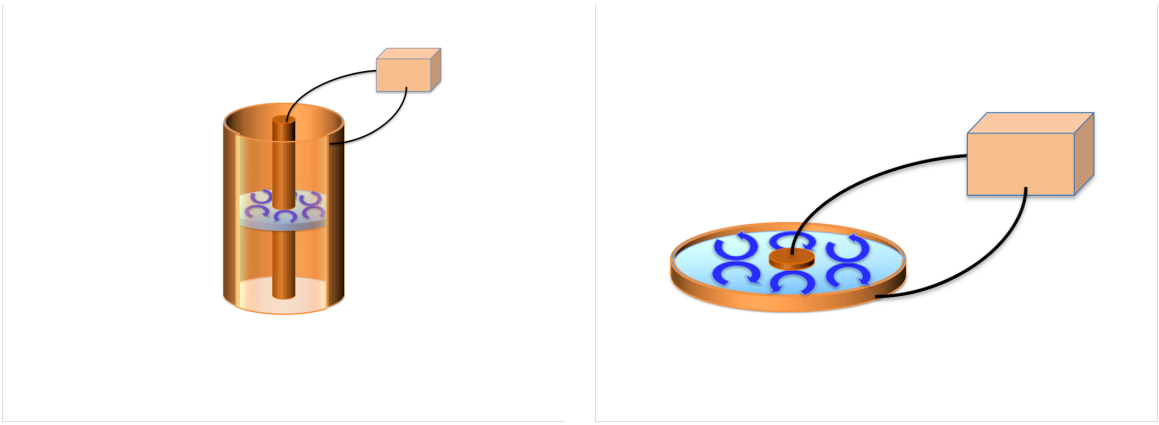
Figure 5.6. Contours of potential temperature θ for inviscid SQG ($\kappa = 0$) at $t = 8,\ 16,\ 26,\ 35$ with initial condition defined by (5.47). Our simulation results (left column) are compared with results reprinted from column (f) of Fig. 1 in [2] (right column).

6. APPLICATION TO THE ELECTROCONVECTION EQUATIONS

Convection in fluid dynamics has been extensively studied by physicsts and mathematicians due to the existence of rigorous mathematical models with numerical simulations supported by precise laboratory measurements. Existing convective systems include Rayleigh-Bénard convection in fluid and fluid mixture confined in rigid plates driven by sufficiently steep temperature gradient [59], Bénard-Marangoni convection driven by the free surface tension [60, 61], Taylor-Couette flow with fluid confined between two rotating cylinders [62], and electrohydrodynamic instability in dielectric liquids [63] and nematic liquid crystals [64, 65].

In this chapter, we examine the electrical convection, *electroconvection* in short, in smectic-A stage liquid crystals. Electroconvection refers to the phenomenon that counter-rotating pairs of vortices appear in a thin layered electrically charged fluid with the application of sufficiently strong electric field as external driving force. For detailed laboratory experiment studies on different geometries, we refer to [66, 67, 68] on rectangular domain, [69] on unsheared annular domain, and [70, 71, 72, 73] on sheared annular domain. In recent years electroconvection has beed applied in bio-technologies [74, 75].

In our work, we follow the experimental settings in [70] and [6] and present an overview of the experiment together with its outcomes in Section 6.1. Our numerical analysis begins with the derivation of the full mathematical model in Section 6.2 followed by its two dimensional model reduction in Section 6.3. We shall see in Section 6.3 that the reduced model contains either the spectral fractional laplacian operator, or the integral fractional laplacian operator. The numerical algorithms for the approximations of the electric potential, the surface charge density and the fluid dynamics are described in Section 6.4. Their general performances together with our numerical assessments of favorable conditions to obtain sustainable convective flow are discussed in Section 6.5. When using the same numerical configurations as in [6, 37], our findings are in good agreement, as reported in Section 6.5.



(a) (b)

Figure 6.1. Experimental settings. The liquid film is located in between two concentric electrodes. (a) the two electrodes extend to infinity (infinite case) and (b) the two electrodes have negligible thickness (slim case). Notice that in both cases, the outer electrode is extended to infinity in the xy-plane (not pictured).

6.1 The physical experiment

The experiment comprises of specially designed apparatus that is enclosed by an aluminum box serving as a Faraday cage to exclude external electromagnetic influences. A weakly conducting, sub-micron thick liquid crystal film is freely suspended between two concentric electrodes; see Figure 6.1 for a schematic description. Several different aspect ratios of the film are investigated. The main experimental control parameter is the imposed DC voltage difference V between the two electrodes. For each aspect ratio there exists a threshold voltage V_c distinguishing the stationary and convective stages of the fluid.

The key objective of the experiment is to study the charge transportation by precise measurement of the current I as the voltage V varies. For $V < V_c$ the fluid is quiescent and charge transportation satisfies the Ohm's law. Once $V > V_c$, the electrical driving force is strong enough to overcome the dissipation and viscosity of the fluid, and the fluid flows in pairs of counter-rotating vortices. The convection carries extra electric current, therefore a significant jump in the slope of the current-voltage curve plot indicates the switch between the stationary and convective stages.

The liquid crystal material

The liquid crystal employed in the experiments is 4-octyl-4-cyanobiphenyl, referred to as 8CB for short. Its chemical formula is given in Figure 6.2. The motion behavior of 8CB is strongly dependent on the temperature, for which 8CB is also classified as of type "thermotropic", in contrast to "lyotropic" liquid crystals. At a temperature lower than $21.5°$C, 8CB is spatially structured as a solid crystal with long-range positional and rotational order. At the temperature higher than $40.5°$C, 8CB arranges spatially randomly as an isotropic fluid with short-range order. Between these two temperatures, 8CB can exhibit short range correlations in some directions but long-range order in others; for instance, the Smectic-A phase and the Nematic phase. Figure 6.2 illustrates the phase sequence of 8CB in response of temperature changes.

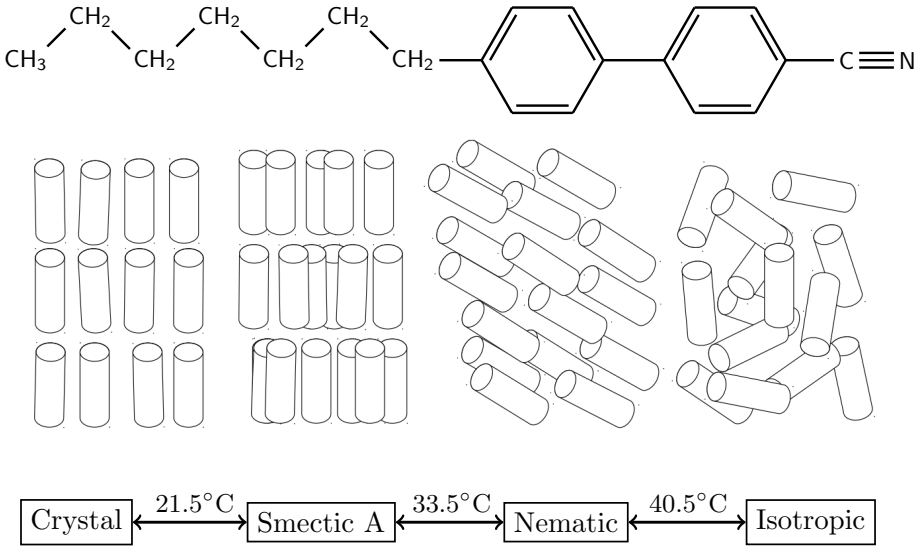


Figure 6.2. Chemical formula of 4-octyl-4-cyanobiphenyl (8CB), see [3]; and phase sequence of 8CB in response to temperature, with figures reprinted from [4].

The experimental temperature is set to $24 \pm 2°$C so that 8CB is in the smectic-A phase, where the crystals are arranged in integer numbered layers with their long axis aligned with the normal of the film plane, as illustrated in Figure 6.2. Under such arrangement the movement of the fluid

molecules across layers is prohibited, thereby imposing a strongly two-dimensional motion within each layer.

In the experiment, the liquid crystal film is freely suspended between two concentric metal electrodes. The film consists of 20-100 smectic layers, each layer being $3.16\,nm$ thick. The width of the annulus is about $1\,cm$. The film is doped with a good electron acceptor to ensure a low but stable ionic conductivity.

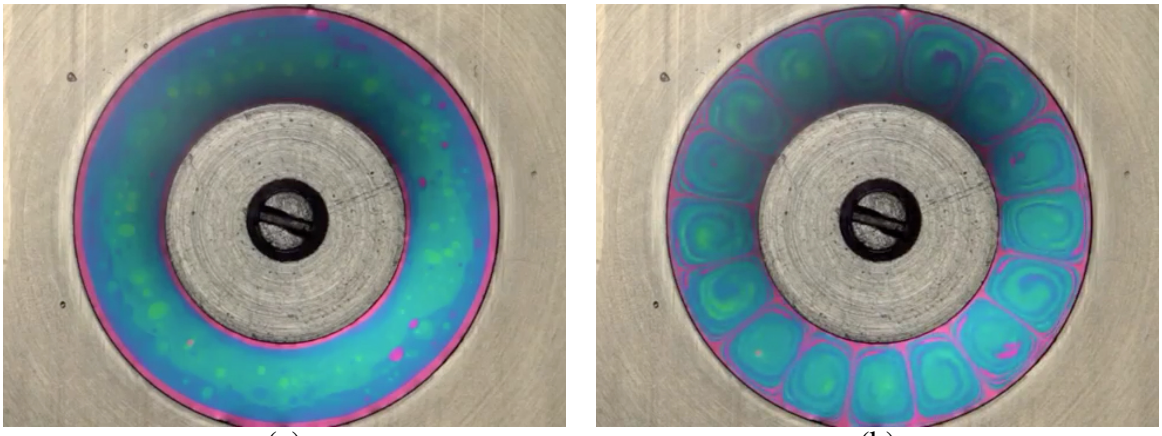
Experimental protocol and outcome

The inner electrode is applied with electric voltage V, whereas the outer electrode is grounded. The applied electric voltage difference drives a low electric current through the film, resulting in an accumulation of the surface charge of the film that positive charges gather near the higher potential region at the inner electrode, while negative charges accumulate close to the grounded outer electrode. Such inverted charge density is unstable and the equilibrium will be broken when the external electric forcing is strong enough. The experimental procedure consists of increasing the voltage V with small positive steps from 0 to above V_c, the critical voltage, and then decreasing V back to 0 with small negative steps. A typical setting is to vary V between 0 and 50 voltes with voltage step 1 volt. Figure 6.3 provides an experimental visualization of the two different states. When V is smaller than V_c, the fluid stays stationary, as shown in Figure 6.3(a); Once V exceeds V_c, the fluid motion is organized into convection vortices in alternating pairs, see Figure 6.3(b). However, the variation in color as shown in the pictures is a consequence of thickness variations, which unfortunately is not an ideal setting for the numerical study. There is also a necessity of conducting a thorough numerical study of electroconvection.

Numerical simplifications

Several simplifications are in order. The electric current inside the fluid is of order 10^{-10} amperes, therefore in our numerical study, the magnetic fields generated by such low current are neglected. This allows us to simplify the Maxwell's equations dramatically; see Subsection 6.2.2.

The fact that the physical thickness of the film ($\approx 0.3\,\mu m$) is much smaller than the width of the



(a) (b)

Figure 6.3. Snapshots of (a) equilibrium state and (b) convective state of the liquid in the physical experiment. The color patterns are reflected by varying the film thickness. Picture reprinted from [5].

annulus ($\approx 1\,cm$) allows us to derive the mathematical models upon neglecting the film thickness and only consider the limiting case of vanishing thickness. The resulting surface charge density conservation relation and the mass and momentum conservation relations for the incompressible liquid are derived on the two dimensional bounded annulus domain Ω.

As to the electrodes, two different settings are considered: the electrodes extend to infinity in the normal direction (referred to as "infinite case") and the electrodes have negligible thickness (referred to as "slim case").

Worth mentioning, the electric field (or electric potential) remains defined in the whole $\mathbb{R}^3$. To fully take advantage of a reduced modeling setting, we resort to equivalent nonlocal representations of the restriction to Ω of the electric potential. Depending on the two cases of electrode configurations, the nonlocal representation involves either the spectral fractional Laplacian (infinite electrodes, Figure 6.1(a)) or the integral fractional Laplacian (slim electrodes, Figure 6.1(b)).

We assess numerically the advantages of the two different electrode configurations. We also provide a numerical study on the effects of three critical nondimensional parameters in the electroconvection process. These three parameters are the Rayleigh number $\mathcal{R}$ representing the ratio

between the electric forcing and viscous dissipation, the Prandtl number $\mathcal{P}$ measuring the ratio between the charge relaxation time and the viscous relaxation time, and the circular electrodes aspect ratio α characterizing the geometry.

6.2 Mathematical models

We describe below the derivation of the electroconvection model from electromagnetic theory and incompressible fluid dynamics. As already mentioned, we consider two settings: infinite and slim electrodes. The former corresponds to the simulation reported in [70, 71, 6] while the later is related to the analysis in [76]. We remark that vanishing charge densities on $\partial\Omega$ are enforced in [76] allowing for smoother charge densities and simplifications of the mathematical model. In this work, we do not make this assumption incompatible with the conservation of charges and refer to Remark 6.1 for additional clarification.

6.2.1 Geometry

We denote by $\Omega := \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ R_i < |\boldsymbol{x}| < R_o\}$, with $0 < R_i < R_o < \infty$ the annular region so that the liquid crystal is confined in the cylindrical domain $\Omega_s := \Omega \times (-s, s)$, where $2s > 0$ stands for the film thickness, and $s \ll d := R_o - R_i$ by the assumption. We also denote by $K_i := \{\boldsymbol{x} \in \mathbb{R}^2 \ : |\boldsymbol{x}| < R_i\}$ and $K_o := \{\boldsymbol{x} \in \mathbb{R}^2 \ : |\boldsymbol{x}| > R_o\}$ the regions in $\mathbb{R}^2$ occupied by the inner and outer electrodes respectively.

In correspondence to the two experimental settings on the electrodes: we have $K_i \times \{0\}$ and $K_o \times \{0\}$ for slim electrodes and $K_i^\infty := K_i \times \mathbb{R}$ and $K_o^\infty := K_o \times \mathbb{R}$ for infinite electrodes. Generically, we use the notations $\mathcal{K}_o$ and $\mathcal{K}_i$ to denote either the slim or infinite electrodes. Furthermore, we use $D_s := \mathbb{R}^3 \setminus (\overline{\Omega}_s \cup \overline{\mathcal{K}}_i \cup \overline{\mathcal{K}}_o)$, with $D := D_0$, to denote the free space.

6.2.2 Electro-magnetism

With the neglect of magnetic effects, the electric field $\boldsymbol{E} : \mathbb{R}^3 \to \mathbb{R}^3$ satisfies $\nabla \times \boldsymbol{E} = 0$, which in turn guarantees the existence of a potential function $\Psi : \mathbb{R}^3 \to \mathbb{R}$ satisfying

$$\boldsymbol{E} = -\nabla\Psi.$$

The relation between the potential Ψ and the surface charge density $q : \partial\Omega_s \to \mathbb{R}$ (units: $\mathrm{C\,m^{-2}}$) is derived from the Gauss law. The argument is separated into two cases depending on the region in $\mathbb{R}^3$. We begin with the argument in the free space. Because there is no charge in the free space D_s, it directly implies that $\nabla \cdot \boldsymbol{E} = 0$, or in terms of potential Ψ,

$$\Delta\Psi = 0 \qquad \text{in } D_s. \tag{6.1}$$

In addition, in accordance to the voltage settings in the experiment, the electric potential Ψ is confined to the appropriate voltages on the two electrodes

$$\Psi = V \quad \text{on} \quad \mathcal{K}_i, \qquad \text{and} \qquad \Psi = 0 \quad \text{on} \quad \mathcal{K}_o. \tag{6.2}$$

In the case of slim electrodes, one needs to add a decay condition at infinity

$$\lim_{|z|\to+\infty} |\Psi(x, y, z)| = 0, \quad \forall (x, y) \in \mathbb{R}^2, \tag{6.3}$$

to close the system.

For the liquid region Ω_s, we recall that even with low conductivity, the charges are located at the surface of Ω_s. As a standard argument, we introduce the "Gaussian pillbox" (see Figure 6.4 and also [77])

$$G(\bar{x}, \bar{y}, r) := (\bar{x} - r, \bar{x} + r) \times (\bar{y} - r, \bar{y} + r) \times (-s - r, s + r)$$

centered at $(\bar{x}, \bar{y}, 0) \in \Omega$ with $r = s$. Applying the Gauss law on $G(\bar{x}, \bar{y}, r)$ yields

$$\varepsilon_0(\boldsymbol{E}(\bar{x}, \bar{y}, s) - \boldsymbol{E}(\bar{x}, \bar{y}, -s)) \cdot \boldsymbol{e}_3 = 2q + O(s). \tag{6.4}$$

Here $\boldsymbol{e}_3 = (0, 0, 1)$ and ε_0 is the permittivity of the free space. Notice that the constant 2 is due to the contributions from the top and bottom sides of the film. Also, the permittivity of the liquid

crystal is assumed to be isotropic in the xy directions and therefore the contributions from the other sides are of order $O(s)$ in (6.4).

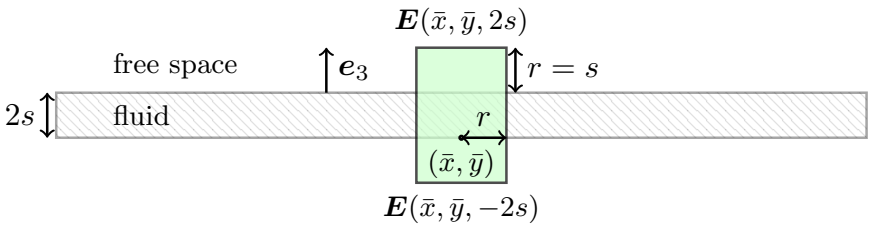


Figure 6.4. Gaussian pillbox enclosing a small region of the fluid.

6.2.3 Fluid dynamics

We use the incompressible Navier-Stokes equations to model the relation between the fluid pressure $p(t) : \Omega_s \to \mathbb{R}$ and the fluid velocity $\boldsymbol{u}(t) : \Omega_s \to \mathbb{R}^3$

$$\rho(\frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla)\boldsymbol{u} + \eta\Delta\boldsymbol{u} + \nabla p = \boldsymbol{L}, \quad \nabla \cdot \boldsymbol{u} = 0, \quad \text{both in } \Omega_s \times \mathbb{R}_+, \tag{6.5}$$

where ρ is the fluid density, η is the shear viscosity and $\boldsymbol{L} := q\nabla\Psi$ is the Lorentz force induced by the application of the external electric fields on the charged molecules. The velocity satisfies the no-slip conditions $\boldsymbol{u} = 0$ at the electrodes $\partial\Omega_s \cap \partial\mathcal{K}_{i/o}$ and the slip conditions $\boldsymbol{u} \cdot \boldsymbol{e}_3 = 0$ on the top and bottom sides of Ω_s.

6.2.4 Small thickness limiting model

We now consider the limiting model when the thickness s is sent to 0. Relation (6.4) between the electric field $\boldsymbol{E} = -\nabla\Psi$ and the surface charge density q becomes

$$\lim_{z\downarrow 0} \frac{\partial}{\partial z}\Psi - \lim_{z\uparrow 0} \frac{\partial}{\partial z}\Psi = -\frac{2q}{\varepsilon_0} \quad \text{in } \Omega, \tag{6.6}$$

Notice that by symmetry along the xy-plane, relations (6.1), (6.2), and (6.6) immediately reduce to a system of elliptic partial differential equations on the upper half plane $D^+ := D \cap \{z \geq 0\}$. The equations consist of

$$\Delta \Psi = 0 \quad \text{in } D^+, \tag{6.7}$$

together with the boundary conditions

$$\Psi = V \text{ on } \mathcal{K}_i^+, \quad \Psi = 0 \text{ on } \mathcal{K}_o^+, \quad \text{and } \lim_{z \to 0} \partial_z \Psi = -\frac{q}{\varepsilon_0} \text{ on } \Omega, \tag{6.8}$$

where $\mathcal{K}_{i/o}^+ := \mathcal{K}_{i/o} \cap \{z \geq 0\}$.

As indicated by the smectic-A liquid crystal motion property, the fluid only moves in the xy-plane, i.e. $\boldsymbol{u} \cdot \boldsymbol{e}_3 = 0$. Hence, the Navier-Stokes system (6.5) reduces to

$$\rho_\Omega (\frac{\partial}{\partial t} + \boldsymbol{u}_\Omega \cdot \nabla_\Omega) \boldsymbol{u}_\Omega + \eta_\Omega \Delta_\Omega \boldsymbol{u}_\Omega + \nabla_\Omega p_\Omega = \boldsymbol{L}_\Omega, \quad \nabla_\Omega \cdot \boldsymbol{u}_\Omega = 0, \quad \text{both in } \Omega \times \mathbb{R}_+, \tag{6.9}$$

where $p_\Omega(t) := p(t)|_\Omega : \Omega \to \mathbb{R}$, and $\boldsymbol{u}_\Omega(t) : \Omega \to \mathbb{R}^2$ denotes the first two components of $\boldsymbol{u}$ restricted to Ω and ρ_Ω, η_Ω are the two-dimensional fluid mass density (units: $\mathrm{kg\,m^{-2}}$) and two-dimensional fluid shear viscosity (units: $\mathrm{Pa\,s\,m}$) respectively. The external force $\boldsymbol{L}_\Omega$ is the projection of the Lorenz force $\boldsymbol{L}$ on the plane supporting Ω, namely, $\boldsymbol{L}_\Omega = q \nabla_\Omega \Psi_\Omega$. Lastly, the boundary conditions on the velocity become

$$\boldsymbol{u}_\Omega = 0 \qquad \text{on } \partial\Omega \times \mathbb{R}_+. \tag{6.10}$$

Inside the fluid Ω, the charge density flux and electric field are assumed to satisfy the Ohm's law, i.e., across any closed curves with outward pointing normal $\boldsymbol{\nu}$ we have

$$\nabla_\Omega q \cdot \boldsymbol{\nu} = -\sigma_\Omega \boldsymbol{E}_\Omega \cdot \boldsymbol{\nu},$$

where σ_Ω stands for the two-dimensional fluid electrical conductivity (units: S) and $\boldsymbol{E}_\Omega$ denotes the

first two components of $\boldsymbol{E}$ (projection of $\boldsymbol{E}$ onto the plane supporting Ω). With this assumption, the conservation of surface charge density reads

$$\frac{\partial}{\partial t}q + \nabla_\Omega \cdot (\boldsymbol{u}_\Omega q - \sigma_\Omega \boldsymbol{E}_\Omega) = 0 \qquad \text{in } \Omega \times \mathbb{R}_+,$$

or

$$\frac{\partial}{\partial t}q + \boldsymbol{u}_\Omega \cdot \nabla_\Omega q - \sigma_\Omega \Delta_\Omega \Psi_\Omega = 0 \qquad \text{in } \Omega \times \mathbb{R}_+, \tag{6.11}$$

using the electric potential $\Psi_\Omega := \Psi|_\Omega$. The surface charge density q in (6.11) is defined up to a constant fixed by the total charge conservation relation

$$\int_\Omega q(t) = \int_\Omega q(0), \qquad \forall t > 0. \tag{6.12}$$

In summary, the electric potential Ψ, the surface charge density q, and the velocity and pressure $(\boldsymbol{u}_\Omega, p_\Omega)$ are related by the system of differential equations (6.7), (6.9), (6.11), and (6.12), which are supplemented with the boundary conditions (6.8), (6.10) and (6.12). From now on, without ambiguity, we drop the subscript Ω on $\boldsymbol{u}_\Omega$, p_Ω.

6.2.5 Nondimensional model

To sort out the effect of the different parameters present in the model, we follow [37] and rewrite the governing equations (6.7), (6.6), (6.9), and (6.11) using the rescaled variables

$$\hat{\boldsymbol{x}} := \frac{\boldsymbol{x}}{d}, \quad \hat{t} := \frac{\sigma_\Omega}{\varepsilon_0 d}t, \quad \widehat{\Omega} := \{\hat{\boldsymbol{x}} \ : \boldsymbol{x} \in \Omega\}, \quad \text{and} \quad \widehat{D}^+ := \{\hat{\boldsymbol{x}} \ : \boldsymbol{x} \in D^+\},$$

where $d := R_o - R_i$ is the distance between the two electrodes. Similarly, we set $\widehat{\mathcal{K}}^+_{i/o} := \{\hat{\boldsymbol{x}} \ : \boldsymbol{x} \in \mathcal{K}^+_{i/o}\}$ for the electrodes. In addition, we define the rescaled functions

$$\widehat{\Psi}(\hat{\boldsymbol{x}}, \hat{t}) := \frac{1}{V}\Psi(\boldsymbol{x}, t), \quad \widehat{\Psi}_{\widehat{\Omega}} = \widehat{\Psi}|_{\widehat{\Omega}}, \qquad \hat{q}(\hat{\boldsymbol{x}}, \hat{t}) := \frac{d}{\epsilon_0 V}q(\boldsymbol{x}, t),$$
$$\widehat{\boldsymbol{u}}(\hat{\boldsymbol{x}}, \hat{t}) := \frac{\epsilon_0}{\sigma_\Omega}\boldsymbol{u}(\boldsymbol{x}, t), \qquad \hat{p}(\hat{\boldsymbol{x}}, \hat{t}) := \frac{\epsilon_0^2}{\sigma_\Omega^2 \rho_\Omega}p(\boldsymbol{x}, t).$$

To make the expressions more readable, we drop the notation $\hat{\cdot}$ and for now on only consider the rescaled nondimensional variables and functions. The rescaling allow us to have

$$\frac{\partial}{\partial t} q + \boldsymbol{u} \cdot \nabla_\Omega q - \Delta_\Omega \Psi_\Omega = 0 \qquad \text{in } \Omega \times \mathbb{R}_+, \tag{6.13}$$

with $\int_\Omega q(t) = \int_\Omega q(0)$ for the conservation of surface charge density relation,

$$-\Delta \Psi = 0 \quad \text{ in } D^+ \times \mathbb{R}_+, \tag{6.14}$$

$$\Psi = 1 \text{ on } \mathcal{K}_i^+, \quad \Psi = 0 \text{ on } \mathcal{K}_o^+, \quad \text{and } \lim_{z \downarrow 0} \partial_z \Psi = -q \text{ on } \Omega \tag{6.15}$$

(together with $\lim_{z \to +\infty} \Psi = 0$ in the case of slim electrodes) for the relations between the electric potential and the surface charge density, and

$$\frac{\partial}{\partial t} \boldsymbol{u} + (\boldsymbol{u} \cdot \nabla_\Omega) \boldsymbol{u} - \mathcal{P} \Delta_\Omega \boldsymbol{u} + \nabla_\Omega p = -\mathcal{R}\mathcal{P} q \nabla_\Omega \Psi_\Omega, \qquad \nabla_\Omega \cdot \boldsymbol{u} = 0, \tag{6.16}$$

in $\Omega \times \mathbb{R}_+$ with boundary conditions $\boldsymbol{u} = 0$ on $\partial\Omega$ for the Navier-Stokes system.

The two dimensionless parameters $\mathcal{P}$ and $\mathcal{R}$ appearing in (6.16) are the Prandtl and Rayleigh numbers. They are given by

$$\mathcal{P} := \frac{\epsilon_0 \eta_\Omega}{\rho_\Omega \sigma_\Omega d}, \quad \text{ and } \quad \mathcal{R} := \frac{\epsilon_0^2 V^2}{\eta_\Omega \sigma_\Omega}. \tag{6.17}$$

The Prandtl number indicates the fluid viscous relaxation ability relative to its charge relaxation ability while the Rayleigh relates the electric forcing with the dissipation forces. In Section 6.5, we propose a numerical study determining what range of parameters allows for electroconvection. We also introduce a geometric characteristic parameter $\alpha := R_i/R_o \in (0, 1)$ so that

$$R_i = \frac{\alpha}{1-\alpha}, \qquad \text{and} \qquad R_o = \frac{1}{1-\alpha}. \tag{6.18}$$

As we shall see in Section 6.5, this parameter affects the number of vortices during the electrocon-

vection process.

6.3 Model reduction

The electroconvection governing equations (6.13)-(6.16) are mainly two dimensional (i.e., defined on Ω) except for (6.14), where the electric potential must be computed in the entire free space D^+. However, we shall notice that only its trace on Ω is required in (6.13) and (6.16). This, together with the two different electrode configurations (infinite and slim), is exploited in Subsections 6.3.1 and 6.3.2 to replace (6.14) with nonlocal problems in Ω.

6.3.1 Infinite electrodes

In this model we assume that the two electrodes extend to infinity along the z directions, refer to Figure 6.5 for an illustration. This is the setting considered for instance in [76]. However, because we do not impose vanishing charge densities on $\partial\Omega$, our model does not reduce to the one analyzed in [76]. To make it worse, it appears that less regular charge densities are to be expected (see Remark 6.1).

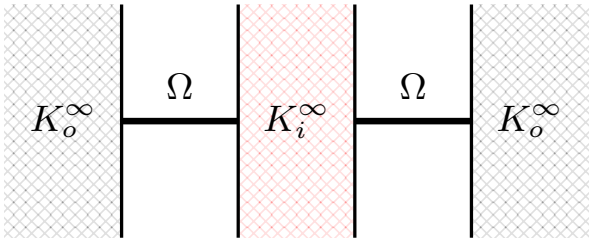


Figure 6.5. Domain and boundaries, cross section view in the infinite electrodes case; compare with Figure 6.6.

We decompose Ψ into two parts $\Psi = \Psi_0 + \Psi_K$. The component Ψ_K accounts for the dimensionless voltages imposed on the electrodes, and is defined as the solution to

$$\Delta\Psi_K = 0 \quad \text{in } D^+ \tag{6.19}$$

together with $\Psi_K = 1$ on $\mathcal{K}_i^+$, $\Psi_K = 0$ on $\mathcal{K}_o^+$ and $\lim_{z\downarrow 0} \partial_z \Psi_K = 0$ on Ω. Notice that Ψ_K is independent of the z variable, and is axis-symmetric for (x, y) variables. In fact, its exact expression is given by

$$\Psi_K(x, y, z) = \eta(x, y) := \frac{\ln(\sqrt{x^2 + y^2}(1 - \alpha))}{\ln(\alpha)}. \tag{6.20}$$

The second part $\Psi_0 = \Psi - \Psi_K$ depends on the charge density q and solves

$$-\Delta\Psi_0 = 0 \quad \text{in } D^+, \quad \Psi_0 = 0 \quad \text{on } \mathcal{K}_i^+ \cup \mathcal{K}_o^+, \quad \text{and} \quad \lim_{z\downarrow 0} \partial_z \Psi_0 = -q \quad \text{on } \Omega.$$

Following the Stinga-Torrea extension [21], we realize that the trace $\Psi_{0,\Omega} := \Psi_0|_\Omega$ satisfies the following non-local partial differential equation on Ω

$$(-\Delta_\Omega)^{\frac{1}{2}} \Psi_{0,\Omega} = q \quad \text{in } \Omega, \tag{6.21}$$

where $(-\Delta_\Omega)^{\frac{1}{2}}$ is the spectral Laplacian defined in (1.1).

Returning to $\Psi = \Psi_0 + \Psi_K$, we find that its restriction to Ω, $\Psi_\Omega := \Psi|_\Omega$ satisfies

$$\Psi_\Omega = \Psi_{0,\Omega} + \eta, \tag{6.22}$$

where the expression of η is given in (6.20). The reduced model system, defined on Ω, consists of (6.13), (6.16), (6.21) and (6.22).

Remark 6.1. We have already mentioned that compared with [76], we do not impose $q|_{\partial\Omega} = 0$ but rather $\int_\Omega q = C$. In particular, the latter implies that the charge density q does not belong to $H_0^1(\Omega)$, the domain of the fractional operator $(-\Delta_\Omega)^{\frac{1}{2}}$. Consequently we are unable to exploit the relation

$$-(\nabla_\Omega \cdot \nabla_\Omega)(-\Delta_\Omega)^{-\frac{1}{2}} = (-\Delta_\Omega)(-\Delta_\Omega)^{-\frac{1}{2}} = (-\Delta_\Omega)^{\frac{1}{2}},$$

to further simplify the charge relation (6.13) as

$$\frac{\partial}{\partial t}q + \boldsymbol{u}\cdot\nabla_\Omega q + (-\Delta_\Omega)^{\frac{1}{2}}q = 0 \quad \text{ in } \Omega\times\mathbb{R}_+.$$

The above simplification is the starting point of the analysis proposed in [76]. We do not make such assumption because it implies

$$\lim_{t\to\infty}\int_\Omega q(t) = 0,$$

which is incompatible with the surface charge density conservation required in our context.

Remark 6.2. The decomposition (6.22) of Ψ_Ω happens to correspond to the definition of the spectral fractional laplacian with non vanishing boundary condition proposed in [78].

6.3.2 Slim electrodes

Instead of assuming infinite electrodes in the z direction, we now consider electrodes with negligible height as illustrated in Figure 6.6.

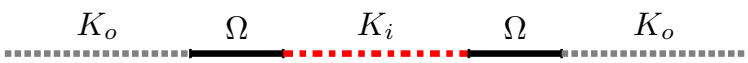


Figure 6.6. Domain and boundaries, cross section view in the slim electrodes case; compare with Figure 6.5.

Same as in the infinite electrodes case, we separate the electrode and charge contributions in the electric potential

$$\Psi = \Psi_0 + \Psi_K. \tag{6.23}$$

In view of (6.14), (6.15) and the decay relation $\lim_{z\to+\infty}\Psi = 0$, the first part Ψ_K is defined as the solution of

$$\Delta\Psi_K = 0, \quad \text{in } \mathbb{R}^2\times\mathbb{R}_+, \qquad \Psi_K = \eta, \quad \text{on } \mathbb{R}^2, \qquad \lim_{z\to+\infty}\Psi_K = 0. \tag{6.24}$$

Here $\eta : \mathbb{R}^2 \to \mathbb{R}$ matches the imposed the electrode voltages and is extended harmonically in Ω:

$$-\Delta_\Omega \eta = 0 \quad \text{in } \Omega, \quad \eta = 0 \quad \text{on } K_o, \quad \text{and} \quad \eta = 1 \quad \text{on } K_i.$$

In fact, it it not difficult to obtain that the exact expression of η on Ω matches (6.20).

The second component $\Psi_0 = \Psi - \Psi_K$ satisfies

$$-\Delta \Psi_0 = 0 \quad \text{in } D^+, \qquad \partial_z \Psi_0 = -q - \partial_z \Psi_K \quad \text{on } \Omega, \qquad \Psi_0 = 0 \quad \text{on } K_i \cup K_o \tag{6.25}$$

together with the decay condition

$$\lim_{z \to +\infty} \Psi_0 = 0. \tag{6.26}$$

It depends on the value of $\partial_z \Psi_K|_\Omega$ we determine now.

When expressed in cylindrical coordinates (r, θ, z), a standard technique based on separation of variables $\Psi_K(r, \theta, z) = X(r, \theta)Z(z)$ (cf. [79, Chapter 4.1]) reveals that the general solution to (6.24) takes the form

$$\Psi_K(r, \theta, z) = \sum_{m=0}^{\infty} \int_0^\infty e^{-kz} J_m(kr)[A_m(k)\cos(m\theta) + B_m(k)\sin(m\theta)]\, dk, \tag{6.27}$$

for $r \geq 0$, $\theta \in [0, 2\pi)$, $z > 0$, and $t \in (0, \mathtt{T})$. Here $J_m(\cdot)$, $m = 0, 1, \cdots$ are Bessel's function of the first kind, while $A_m(k)$ and $B_m(k)$ are coefficient functions that shall be uniquely determined by the boundary condition on the lower boundary $z = 0$. Indeed, since $\Psi_K(r, \theta, 0) = \eta(r)$, a comparison between the coefficients of these two functions reveals that $B_0 = 0$, and $A_m = B_m = 0$ for all $m \geq 1$, therefore

$$\Psi_K(r, \theta, 0) = \eta(r) = \int_0^\infty e^{-kz} J_0(kr) A_0(k)\, dk, \quad \text{for } r \geq 0. \tag{6.28}$$

It remains to recover $A_0(k)$, for which we take advantage of the following orthogonality property

(see [80, Section 11.2])

$$\int_0^\infty J_m(kr)J_m(k'r)rdr = \frac{1}{k}\delta_{k,k'}, \quad \text{for } m = 0, 1, 2, \cdots,$$

where $k > 0$ and $\delta_{k,k'}$ denotes the Kronecker delta. We multiply both ends by $J_0(k'r)r$ in (6.28), and integrate with respect to r to obtain

$$\begin{aligned}\int_0^\infty \eta(r)J_0(k'r)r\ dr &= \int_0^\infty J_0(k'r)r \int_0^\infty J_0(kr)A_0(k)\ dk\ dr \\ &= \int_0^\infty A_0(k) \int_0^\infty J_0(kr)J_0(k'r)r\ dr\ dk \\ &= \int_0^\infty \frac{1}{k}A_0(k)\delta(k-k')\ dk = \frac{1}{k'}A_0(k').\end{aligned}$$

It follows from formula 2.12.28.2 in [81] that for all $k > 0$,

$$\begin{aligned}A_0(k) &= k\int_0^\infty \eta(r)J_0(kr)r\ dr \\ &= \int_0^{R_i} J_0(sk)sds + \frac{1}{\ln\alpha}\int_{R_i}^{R_o}(\ln(1-\alpha)+\ln s)J_0(sk)s\ ds \\ &= \frac{R_i}{k}J_1(R_ik) + \frac{1}{\ln\alpha}\Big[\frac{R_o\ln(1-\alpha)}{k}J_1(r_1k) + \frac{1}{k^2}\Big(J_0(R_ok) + (R_o\ln R_o)kJ_1(R_ok) - 1\Big)\Big] \\ &\quad - \frac{1}{\ln\alpha}\Big[\frac{R_i\ln(1-\alpha)}{k}J_1(R_ik) + \frac{1}{k^2}\Big(J_0(R_ik) + (R_i\ln R_i)kJ_1(R_ik) - 1\Big)\Big] \\ &= \frac{1}{k^2\ln\alpha}\Big(J_0(R_ok) - J_0(R_ik)\Big).\end{aligned}$$

Upon determining all coefficients A_m and B_m, $m = 0, 1, 2, \cdots$, we conclude from (6.27) that for $r \in (R_i, R_o)$ and $\theta \in [0, 2\pi)$,

$$\begin{aligned}\lim_{z\downarrow 0}\partial_z\Psi_K(r,\theta,z,t) &= -\int_0^\infty kJ_0(kr)A_0(k)dk = -\frac{1}{\ln\alpha}\int_0^\infty (J_0(R_ok) - J_0(R_ik))J_0(rk)dk \\ &= -\frac{1}{\ln\alpha}\left(\frac{1}{R_o}{}_2F_1(\frac{1}{2},\frac{1}{2};1;\frac{r^2}{R_o^2}) - \frac{1}{r}{}_2F_1(\frac{1}{2},\frac{1}{2};1;\frac{R_i^2}{r^2})\right),\end{aligned}$$

where in the last equality we applied Formula 6.574.1 in [82]. Here the function ${}_2F_1$ denotes the

hyper-geometric function, see [83, Chapter 15]. For notational convenience, we write

$$\lim_{z\downarrow 0} \partial_z \Psi_K = g, \quad \text{in } \Omega,$$

with

$$g(r) := -\frac{1}{\ln \alpha}\left(\frac{1}{R_o}{}_2F_1(\frac{1}{2},\frac{1}{2};1;\frac{r^2}{R_o^2}) - \frac{1}{r}{}_2F_1(\frac{1}{2},\frac{1}{2};1;\frac{R_i^2}{r^2})\right). \tag{6.29}$$

Remark 6.3. Notice that for given constants $a,\, b > 0$ the hypergeometric function ${}_2F_1(a,b;a+b;x)$ possesses the following asymptotic property

$$\lim_{x\to 1^-} \frac{{}_2F_1(a,b;a+b;x)}{-\ln(1-x)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}.$$

It follows that even though $g(r)$ is singular at the boundaries of Ω, it is still $L^2(\Omega)$ integrable. Consequently, the system (6.25) for Ψ_0 can be uniquely determined by $q \in L^2(\Omega)$ together with g given by (6.29).

For Ψ_0 determined by (6.25), we denote by $\Psi_{0,\Omega}$ to be the trace of Ψ_0 on Ω, and define $\widetilde{\Psi}_{0,\Omega}$ to be the zero extension of $\Psi_{0,\Omega}$ to $\mathbb{R}^2$, i.e.,

$$\widetilde{\Psi}_{0,\Omega} = \Psi_0 \qquad \text{on } \Omega, \qquad \text{and} \qquad \widetilde{\Psi}_{\Omega} = 0 \quad \text{on } \mathbb{R}^2 \setminus \Omega.$$

It turns out that $\widetilde{\Psi}_{0,\Omega}$ satisfies the following nonlocal problem on $\mathbb{R}^2$

$$(-\Delta_{\mathscr{F}})^{\frac{1}{2}} \widetilde{\Psi}_{0,\Omega} = q + g \qquad \text{in } \Omega, \tag{6.30}$$

where the fractional operator $(-\Delta_{\mathscr{F}})^{\frac{1}{2}}$ is the integral fractional laplacian defined in (1.2). Equivalence between (6.25) and (6.30) is proved in [20] with the slight difference that $\partial_z \Psi_0$ is given on $\mathbb{R}^2$ instead of only in Ω. However, the proposed technique based on comparing the energy functionals that the two functions Ψ_0 and $\widehat{\Psi}$ are unique minimizers extends readily in the present context. The proof is rather technical and relatively independent of the topics of this chapter, therefore is given

in Appendix A.

Combining the relations obtained for Ψ_K and Ψ_0, we conclude that the restriction of $\Psi_\Omega := \Psi|_\Omega$ is given by

$$\Psi_\Omega = \Psi_{0,\Omega} + \eta \tag{6.31}$$

where $\Psi_{0,\Omega}$ satisfies (6.30) with g as in (6.29) and η is given by (6.20). The reduced model for the slim electrodes case defined on Ω consists of equations (6.13), (6.16) and (6.31).

6.4 Numerical algorithms

In this section we propose the numerical algorithms advocated to approximate the fluid dynamic (6.16), the surface charge density convection (6.13), and the two non-local problems for the electric field (6.22) or (6.31) depending on the assumption made on the electrodes. In fact, the time marching algorithm proposed consists of three sub-steps. First, the electric potential is approximated in Ω using the (previous) surface charge density. Second, the surface charge density approximation is updated using the electric potential and (previous) fluid velocity. Third, the fluid velocity (and pressure) is updated with a Lorentz force computed using the surface charge density and electric potential. We describe each step separately below. We remark that the initial surface charge density and velocity are supplied allowing the algorithm to start.

6.4.1 Approximation of the electric potential

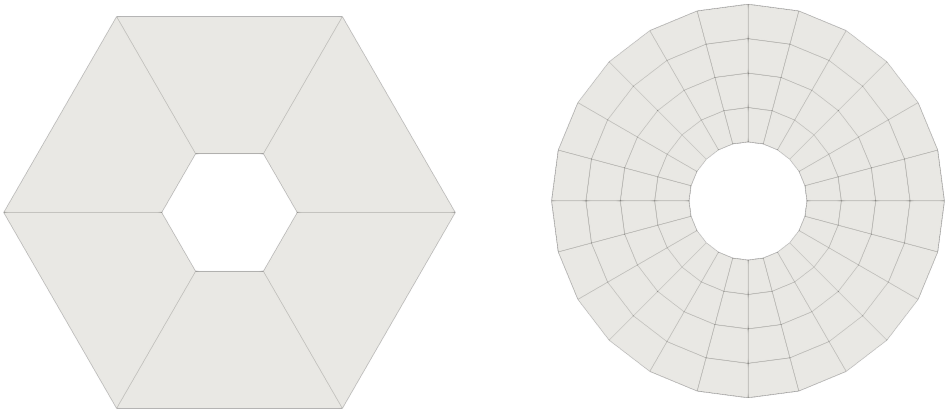
The approximation to the electric potential requires solving (6.21) or (6.30). We now discuss the approximation schemes correspond to the two different configurations separately.

Infinite electrodes

We start with the simpler case of infinite electrodes, where the electric potential satisfies (6.21) involving the spectral Laplacian. Recall that $\Psi_\Omega = \Psi_{0,\Omega} + \eta$, with η given by (6.20). We adopt the numerical procedure consists of sinc quadrature and finite elements, as described in Section 2.7 for the approximation of $\Psi_{0,\Omega}$.

We first construct the non-conforming approximation Ω_h to the annulus domain Ω consists of

subdivisions $\mathcal{T}_h$. The procedure starts with a coarse polygonal approximation of Ω subdivided into quadrilaterals as shown in Figure 6.7(a). This polygonal approximation is then uniformly refined using quad-refinement by connecting midpoints between opposite edges, but placing the boundary vertices on the exact boundary of Ω. This gives rise to a polygonal domain Ω_h and a uniform partition $\mathcal{T}_h$ made of quadrilaterals without hanging nodes, see for instance Figure 6.7(b). Here h denotes the maximum diameter of elements on $\mathcal{T}_h$.



(a) (b)

Figure 6.7. Polygonal approximation Ω_h of Ω for an aspect ratio $\alpha = 0.33$. (a) The coarse initial subdivision; (b) Refined approximation resulted from two successive uniform refinements and placing the boundary vertices on the boundary of Ω.

The number of quadrilaterals in the coarse subdivision depends on the aspect ratio in order to maximize the quality of the subdivision, see the `GridGenerator::hyper_shell` documentation in [84]. Furthermore, the final resolution h required for the simulations presented in Section 6.5 depends on the aspect ratio α as well. Indeed, we shall see in Subsection 6.5.3 that larger aspect ratios result in more pairs of vortices during electroconvection, thereby higher the resolution is a necessity. The number of uniform refinements performed on the coarse subdivisions for each aspect ratio α is listed in Table 6.1.

α	0.1	0.2	0.33	0.452	0.56	0.6446	0.9
$\#\mathcal{T}$	4	5	6	9	12	15	29
# uniform ref.	5	5	5	5	5	5	6

Table 6.1. Infinite electrodes mesh configuration: number of quadrilaterals $\#\mathcal{T}$ and subsequent uniform refinements used in the coarse subdivision for various aspect ratios α.

With $\mathcal{T}_h$ and the mapping $F_T : [0,1]^2 \to T$ defined by (5.39), we obtain the finite element space

$$V_h := \{v \in C^0(\bar{\Omega}_h) \ : \ v|_T \circ F_T \text{ is bilinear for all } T \in \mathcal{T}_h, \text{ and } v|_{\partial\Omega_h} = 0\}.$$

The resulting finite element approximation of $\Psi_{0,\Omega}$ is given by

$$\Psi_{0,h} := \Psi_{0,h}^k := \frac{2k}{\pi} \sum_{j=-N_-}^{N_+} e^{s_j} \Phi_h(s_j; q) \in V_h, \tag{6.32}$$

where $\Phi_h(s_j; q) \in V_h$ solves

$$\int_\Omega \Phi_h(s_j; q)\varphi_h + e^{2s_j} \int_\Omega \nabla\Phi_h(s_j; q) \cdot \nabla\varphi_h = \int_\Omega q\varphi_h, \qquad \text{for all } \varphi_h \in V_h. \tag{6.33}$$

In the numerical simulations proposed in Section 6.5, we fix sinc quadrature stepping $k = 0.04$, allowing $N_- = N_+ = 62$.

Returning to the decomposition (6.22), we achieve an approximation

$$\Psi_h := \Psi_{0,h} + \pi_h\eta \in V_h, \tag{6.34}$$

where π_h stands for the L^2 projection onto V_h.

Slim electrodes

We now consider the case of slim electrodes. In this case, the electric potential on Ω is given by $\Psi_\Omega = \Psi_{0,\Omega} + \eta$ where $\Psi_{0,\Omega}$ satisfies (6.30). Due to the existence of integral fractional laplacian,

the approximation scheme summarized in Section 2.8 will be advocated.

We begin with providing an explicit expression of the truncated domain $\Omega^M(s)$ from recalling that

$$\Omega = \left\{ (x,y) \in \mathbb{R}^2 \ : \ \frac{\alpha}{1-\alpha} < \sqrt{x^2+y^2} < \frac{1}{1-\alpha} \right\},$$

therefore, we define

$$\Omega^M(s) := \begin{cases} \{(1+e^{-\frac{s}{2}}(M+1))(x,y) \ : \ \sqrt{x^2+y^2} < 1/(1-\alpha)\}, & e^{-\frac{s}{2}} \geq 1, \\ \{(M+2)(x,y) \ : \ \sqrt{x^2+y^2} < 1/(1-\alpha)\}, & e^{-\frac{s}{2}} < 1, \end{cases}$$

where $M > 1$ is a given truncation parameter.

We postpone for the moment the discussion regarding the automatic subdivision of $\Omega^M(s)$ but denote by $\Omega_h^M(s)$ its polygonal approximation and by $V_h^M(s)$ the associated finite element space based on continuous piecewise bilinear finite elements vanishing on $\partial\Omega_h^M(s)$. The approximation $\Psi_{0,h} \in V_h$ of $\Psi_{0,\Omega}$ satisfying (6.30) is then given by the relations

$$\frac{k}{\pi} \sum_{j=-N_-}^{N_+} e^{\frac{s_j}{2}} \int_{\Omega_h} (\Phi_h(s_j;\Psi_{0,h}) + \Psi_{0,h})\varphi_h = \int_{\Omega_h} (q+g)\varphi_h, \qquad \forall \varphi_h \in V_h,$$

where $\Phi_h(s_j;\Psi_{0,h}) := \Phi_h^{k,M}(s_j;\Psi_{0,h}) \in V_h^M(s_j)$ solves

$$e^{s_j} \int_{\Omega_h^M(s_j)} \Phi_h(s_j;\Psi_{0,h})\varphi_h + \int_{\Omega_h^M(s_j)} \nabla\Phi_h(s_j;\Psi_{0,h}) \cdot \nabla\varphi_h = -e^{s_j} \int_{\Omega_h} \Psi_{0,h}\varphi_h,$$

for all $\varphi_h \in V_h^M(s_j)$. For the numerical experiments provided in Section 6.5, we take $k = 0.04$, $N_- = 62$, $N_+ = 124$ and $M = 3$.

Returning to the decomposition (6.31), we define

$$\Psi_h := \Psi_{0,h} + \pi_h\eta \in V_h, \tag{6.35}$$

where η is given by (6.20).

We devote the remaining of this subsection on the description of the automatic generation of polygonal approximation $\Omega_h^M(s)$ to $\Omega^M(s)$ and its associated subdivisions. We denote by $R^M(s)$ the radius of the truncated circular domain $\Omega^M(s)$. We start with a coarse subdivision made of 25 quadrilaterals as illustrated in Figure 6.8(a). To ensure the accuracy while keeping the complexity of the resulting linear system under control, for a given $0 < h \leq 1$, the refinement procedure consists of

- Refine the fluid domain Ω_h uniformly until all quadrilaterals have diameters no larger than h, with possible additional refinements on $\Omega_h^M(s)$ outside Ω_h to keep the number of hanging nodes to a maximum of one on each edge. During this refinement process, newly created vertices at the boundary of Ω_h are placed on the boundary of Ω.

- An exponential grading is performed outside the convex hull of Ω_h, i.e., vertices on the azimuthal directions are placed at radii r_j with

$$r_j := e^{jh_0}/(1-\alpha), \quad j = 1, 2, \cdots, \lceil M/h \rceil \quad \text{where} \quad h_0 = \ln(R^M(s)(1-\alpha))h/M;$$

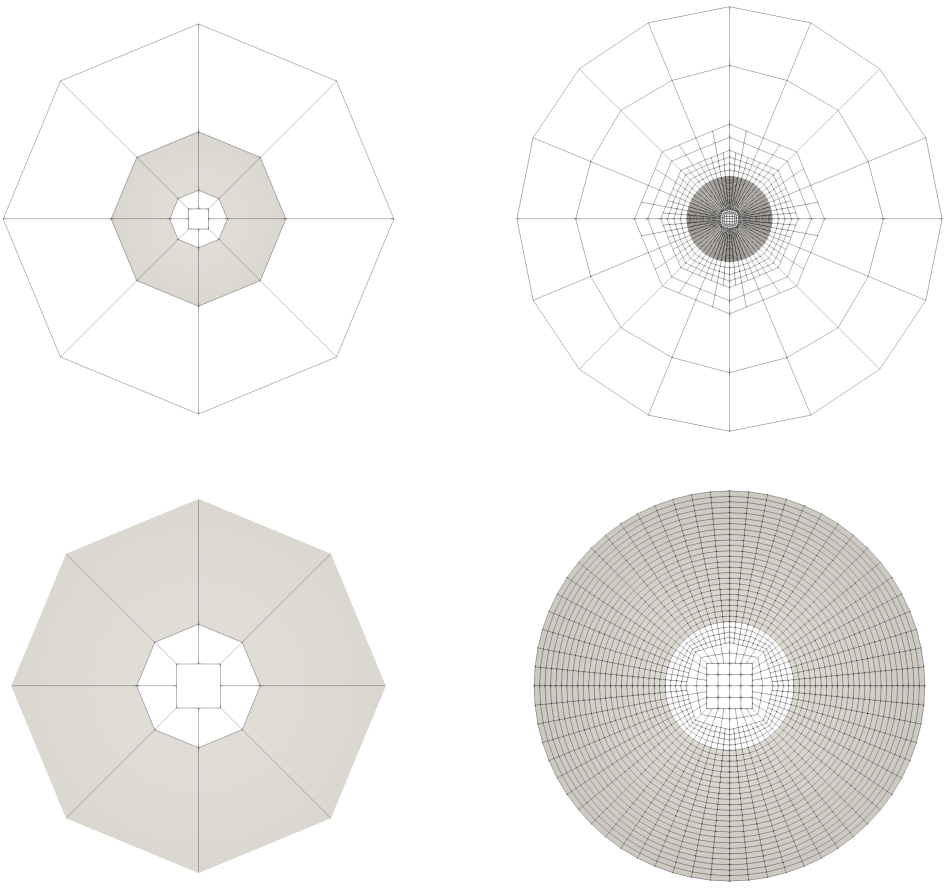
We refer to Figure 6.8(b) for an illustration.

Similar to the case of infinite elecrodes configuration, the final resolution depends on the aspect ratio α to accommodate for the number of vortices during electroconvection. In the simulations presented in Section 6.5, we performed 5 uniform refinements for $\alpha \in \{0.1, 0.2, 0.33, 0.452\}$, 6 for $\alpha \in \{0.56, 0.6446\}$ and 7 for $\alpha = 0.8$.

6.4.2 Approximation of the charge density

We now discuss the approximation of the surface charge density satisfying (6.13). Because Ψ_K is harmonic in Ω in both electrodes configurations, we have

$$-\Delta_\Omega \Psi_\Omega = -\Delta_\Omega \Psi_{0,\Omega}.$$



(a) (b)

Figure 6.8. Polygonal approximation $\Omega_h^M(s)$ (first row) and zoom of the liquid region Ω_h (second row) with $M = 3$ and $s = 1$ associated to Ω of aspect ratio $\alpha = 0.33$. The approximation Ω_h of the liquid domain Ω is in gray. Column (a) Initial subdivision and Column (b) Three successive iterations of the refinement procedure.

Substituting back to (6.13), the surface charge density conservation relation reduces to

$$\frac{\partial}{\partial t} q + \boldsymbol{u} \cdot \nabla_\Omega q - \Delta_\Omega \Psi_{0,\Omega} = 0 \qquad \text{in } \Omega \times \mathbb{R}_+.$$

Notice that for a given electric potential $\Psi_{0,\Omega}$, the above partial differential equation for q is a standard transport equation. It is approximated with an explicit Runge-Kutta 2 method in time and standard finite elements in space.

We recall that $\mathcal{T}_h$ denotes a partition of a polygonal approximation Ω_h of the liquid domain Ω (see Subsection 6.4.1). We introduce the finite element space for the surface charge density

$$Q_h := \left\{ v \in C^0(\overline{\Omega}_h) \ : \ v|_T \circ F_T \text{ is bilinear for all } T \in \mathcal{T}_h, \quad \text{and } \int_{\Omega_h} v = 0 \right\}.$$

Hence, given an approximation $\Phi_{0,h} \in V_h$ of $\Phi_{0,\Omega}$ determined as discussed in Subsection 6.4.1, the approximation $q_h(t) \in Q_h$ of $q(t)$ is defined as the solution to

$$\begin{aligned} \int_{\Omega_h} \frac{\partial}{\partial t} q_h(t)\varphi_h + \int_{\Omega_h} \boldsymbol{u}_h(t) \cdot \nabla_\Omega q_h(t)\varphi_h \\ + \int_{\Omega_h} \nabla_\Omega \Psi_{0,h} \cdot \nabla_\Omega \varphi_h - \int_{\partial\Omega_h} \nabla_\Omega \Psi_{0,h} \cdot \boldsymbol{\nu}_h \varphi_h = 0, \qquad \forall \varphi_h \in Q_h, \end{aligned} \tag{6.36}$$

where $\boldsymbol{\nu}_h$ is the outward pointing normal to Ω_h (defined almost everywhere). This relation is obtained upon multiplying the surface charge density equation with a test function $\varphi_h \in V_h$ and integrating by parts the electric potential term. We suplement (6.36) with an approximated initial condition $\pi_h q(x, y, 0)$, where π_h denotes the $L^2(\Omega)$ projection onto Q_h.

The SSP-RK2 scheme is advocated for the time discretization. Set $q_h^0 = \pi_h q(0)$ the $L^2(\Omega)$ projection of a given surface charge density onto Q_h and $t_n := n\tau$ for $n = 0, 1, 2, \cdots$ for a time step parameter $\tau > 0$. Given approximation $q_h^n \in V_h$ (from previous time step), $\Psi_{0,h}^n \in V_h$ (obtained from solving (6.34) or (6.35) with q replaced by q_h^n) and $\boldsymbol{u}_h^n$ (from previous time step, see also Subsection 6.4.3) of the charge density $q(t_n)$, electric potential $\Psi(t_n)$ and fluid velocity $\boldsymbol{u}(t_n)$ respectively, we approximate $q(t_{n+1})$ by $q_h^{n+1} \in Q_h$ with two stages as follows. In the first stage we seek $\mu_h^{(1)} \in Q_h$ satisfying

$$\begin{aligned} \int_{\Omega_h} \mu_h^{(1)} \varphi_h = \int_{\Omega_h} q_h^n \varphi_h - \tau \int_{\Omega_h} \boldsymbol{u}_h^n \cdot \nabla_\Omega q_h^n \varphi_h \\ - \tau \int_{\Omega_h} \nabla_\Omega \Psi_{0,h} \cdot \nabla_\Omega \varphi_h + \tau \int_{\partial\Omega_h} \nabla_\Omega \Psi_{0,h} \cdot \boldsymbol{\nu}_h \varphi_h, \end{aligned} \tag{6.37}$$

for all $\varphi_h \in Q_h$. In the second stage, we find $\mu_h^{(2)} \in Q_h$ solving (6.37) but with q_h^n replaced by

$\mu_h^{(1)}$. Then, we set

$$q_h^{n+1} := \frac{1}{2}(q_h^n + \mu_h^{(2)}). \tag{6.38}$$

For the same reason as in Subsection 5.3, in order to eliminate the spurious oscillations introduced from the continuous finite element approximations, we include the smoothness-based second-order maximum principle preserving viscosity method proposed in [36]. The numerical parameters used in the artificial viscosity are those recommended in (5.4) in [57].

6.4.3 Approximation of the fluid dynamic

We discretize the fluid dynamic using backward differentiation scheme of order 2 (BDF-2) coupled with Taylor-Hood finite element approximations for the space discretization; see [85]. Given a subdivision $\mathcal{T}_h$ of Ω_h constructed as in Subsection 6.4.1, the finite element spaces for the velocity and pressure are defined by

$$\begin{aligned}
\boldsymbol{W}_h &:= \Big\{ \boldsymbol{v} \in C^0(\overline{\Omega}_h)^2 \ : \ \boldsymbol{v}|_T \circ F_T \in \mathcal{Q}_2^2, \ \forall T \in \mathcal{T}_h, \ \boldsymbol{v}|_{\partial\Omega_h} = 0 \Big\}, \\
X_h &:= \Big\{ \theta \in C(\overline{\Omega}_h) : \int_{\Omega_h} \theta = 0, \ \theta|_T \circ F_T \in \mathcal{Q}_1, \forall T \in \mathcal{T}_h \Big\},
\end{aligned}$$

where $\mathcal{Q}^i$, $i = 1, 2$, stands for the space of polynomial of (total) degree i.

We start with $\boldsymbol{u}_h^0 := \pi_h \boldsymbol{u}(0)$, the $L^2(\Omega)$ projection of a given initial velocity $\boldsymbol{u}(0)$ onto $\boldsymbol{W}_h$. We assume that at time $t = t_n := n\tau$ (for the same time stepping parameter τ used for the surface charge density approximation in Subsection 6.4.2), we have obtained $\Psi_h^n \in V_h$ given by (6.34) or (6.35) with q replaced by q_h^n and $q_h^{n+1} \in Q_h$ given by (6.38). The approximation $(\boldsymbol{u}_h^{n+1}, p_h^{n+1}) \in \boldsymbol{W}_h \times X_h$ of $(\boldsymbol{u}(t_{n+1}), p(t_{n+1}))$, with $t_{n+1} := (n+1)\tau$, is then defined as satisfying

$$\begin{aligned}
\int_{\Omega_h} \boldsymbol{u}_h^{n+1} \cdot \boldsymbol{v}_h + \frac{2}{3}\tau\mathcal{P} \int_{\Omega_h} \nabla \boldsymbol{u}_h^{n+1} \cdot \nabla \boldsymbol{v}_h - \frac{2}{3}\tau \int_{\Omega_h} \nabla \cdot \boldsymbol{v}_h p_h^{n+1} \\
= \int_{\Omega_h} \left(\frac{4}{3}\boldsymbol{u}_h^n - \frac{1}{3}\boldsymbol{u}_h^{n-1} \right) \cdot \boldsymbol{v}_h - \frac{2}{3}\tau \int_{\Omega_h} \boldsymbol{v}_h \cdot (\boldsymbol{u}_h^n \cdot \nabla)\boldsymbol{u}_h^n \\
- \frac{2}{3}\tau\mathcal{R}\mathcal{P} \int_{\Omega_h} q_n^{n+1} \nabla_\Omega \Psi_h^n \cdot \boldsymbol{v}_h,
\end{aligned} \tag{6.39}$$

and

$$\int_{\Omega_h} \nabla \cdot \boldsymbol{u}_h^{n+1} \theta_h = 0$$

for all $(\boldsymbol{v}_h, \theta_h) \in \boldsymbol{W}_h \times X_h$.

6.5 Numerical simulations

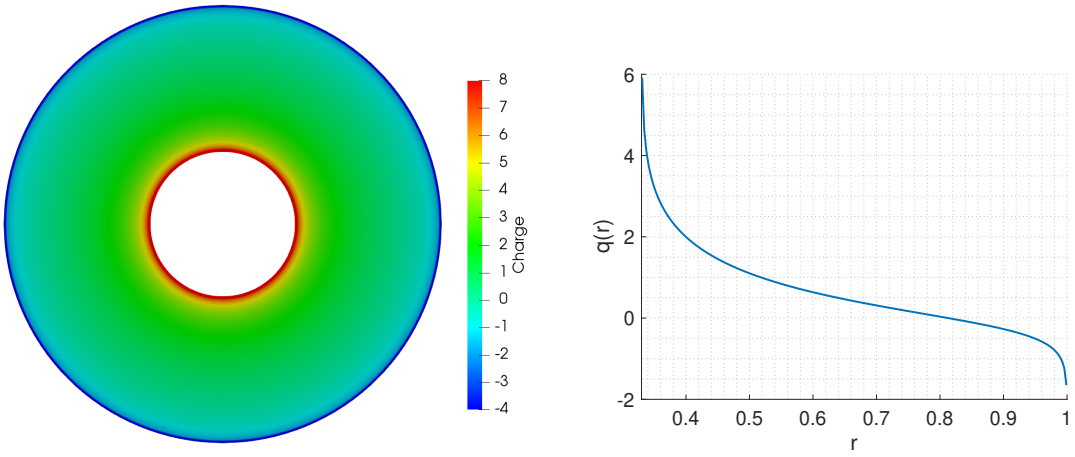
In this section we discuss the numerical results for two different settings with a particular emphasis on the effects of the Prandtl number $\mathcal{P}$, the Rayleigh number $\mathcal{R}$, and the aspect ratio α.

The initial setting is common to all experiments. The fluid is always starting at rest and, for the slim electrode setting, we start with an initial surface charge density in equilibrium with the electric potential as described now. The initial charge density q_0 is chosen so that $\Psi = \Psi_K$ (i.e. $\Psi_{0,\Omega} = 0$) in the decomposition (6.23) for the slim electrodes configuration. This corresponds to setting $q_0 = -g$, where g is given by (6.29). We provide in Figure 6.9 an illustration of the initial surface charge density for $\alpha = 0.33$. Notice that this configuration is unstable as the electric charges are aggregated near the outer boundary where the voltage is minimal. In the simulation below, we break the symmetry by adding a Gaussian white noise of magnitude no larger than 10^{-4}. For comparison, we consider the same initial surface charge density in the infinite electrode case.

In order to detect when and whether electroconvection occurs, we monitor two quantities: the kinetic energy and the circulation energy

$$E_k := \frac{1}{2} \int_\Omega |\boldsymbol{u}|^2 \, d\boldsymbol{x}, \qquad E_{curl} := \int_\Omega |\nabla \times \boldsymbol{u}|^2 \, d\boldsymbol{x}.$$

Because of the imposed white noise to the initial surface charge density, the two quantities E_k and E_{curl} evolves initially. We declare the electroconvection phenomena to occur when both E_k and E_{curl} undergo a relative change greater than 0.1% compared to their respective initial values before time $t = 40$ (we ran our simulation further and did not observe any changes after that). We define the critical Rayleigh number $\mathcal{R}_c$ as the threshold value for which fluids with lower Rayleigh numbers $\mathcal{R} < \mathcal{R}_c$ electroconvection do not occur (in that case the Lorentz force is not strong enough to overcome the electric and viscous dissipation). For Rayleigh numbers above



(a) (b)

Figure 6.9. The initial condition of charge density $q_0 = -g$ with g defined by (6.29) for $\alpha = 0.33$. (a) Initial surface charge density distribution; (b) Cross-section view of the initial surface charge density with $r = \sqrt{x^2 + y^2}$.

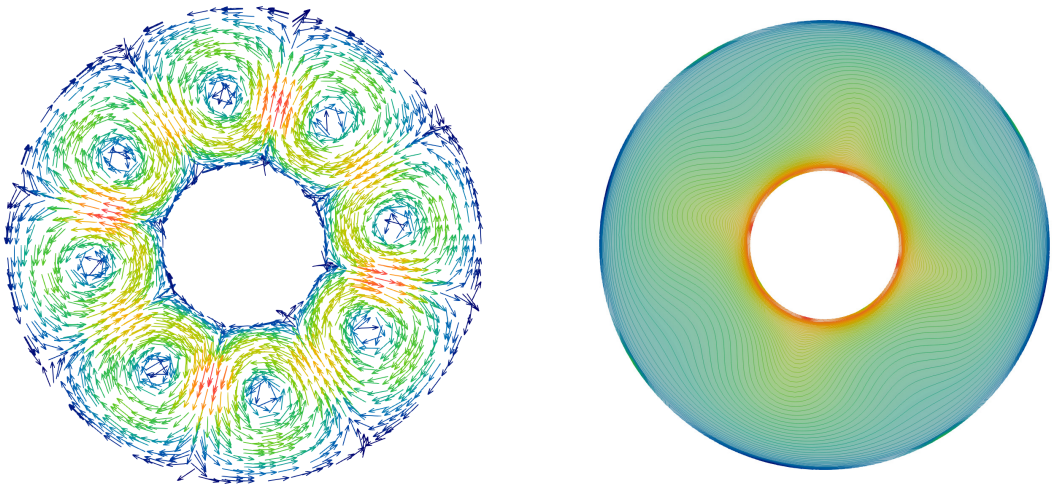
$\mathcal{R}_c$, the axis-symmetric distribution of charge density is broken and vortices appear. The critical vortex pair number $\mathcal{N}_c$ is the number of pairs of vortices during electroconvection at $\mathcal{R} = \mathcal{R}_c$. As an illustration, we provide in Figures 6.10 (where $\alpha = 0.33$) and 6.11 (where $\alpha = 0.56$) the numerical approximation of sustained electroconvection with $\mathcal{N}_c = 4$ and $\mathcal{N}_c = 8$ respectively. Similar structures but with different $\mathcal{N}_c$ are observed for other different aspect ratios.

We summarize the numerical parameters using in all the simulation below in Table 6.2 for the infinite electrodes configuration and in Table 6.3 for the slim electrodes configuration.

6.5.1 Comparison between the infinite and slim electrodes models

For this comparison, we set the aspect ratio to $\alpha = 0.33$ and the Prandtl number to $\mathcal{P} = 10$. When the Rayleigh number is $\mathcal{R} = 100$, electroconvection is observed in the slim electrodes configuration but not in the infinite electrodes configuration, which seems to require more energy to trigger electroconvection. In fact, electroconvection is observed in the infinite electrodes configuration for $\mathcal{R} \geq 250$, see kinetic and circulation energies in Figure 6.12.

Moreover, we find that even when electroconvection occurs in the infinite electrodes configu-



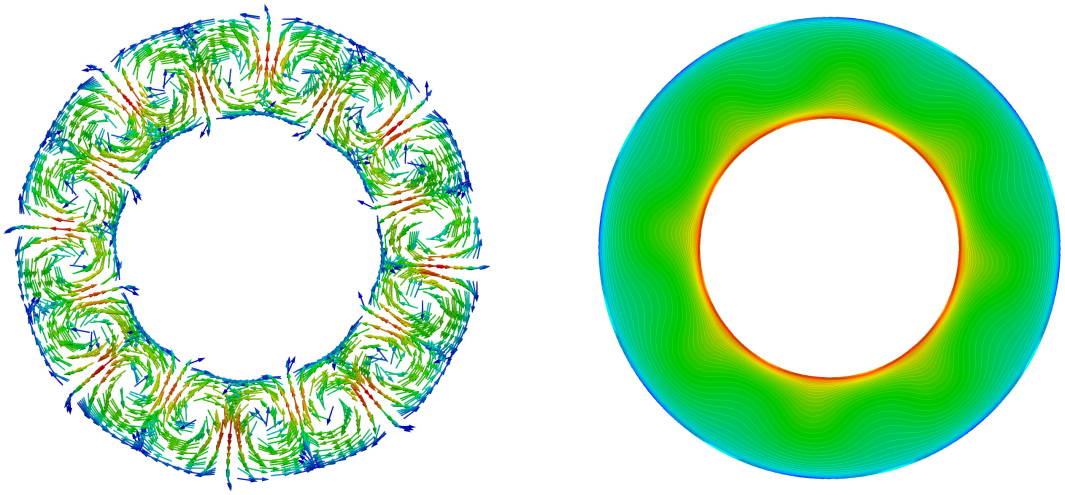
(a) (b)

Figure 6.10. Electroconvection for $\mathcal{P} = 10$, $\mathcal{R} = 100$, and $\alpha = 0.33$ at time $t = 40$. (a) Numerical approximation of the velocity field $\boldsymbol{u}$; (b) Numerical approximation of the electric surface charge density distribution q.

α	τ	DoFs for $\Psi_{0,\Omega}$	DoFs for q	DoFs for $\boldsymbol{u}$ and p
0.1	0.001	4,224	4,224	33,280/4,224
0.2	0.001	5,280	5,280	41,600/5,280
0.33	0.001	7,392	7,392	58,240/7,392
0.452	0.001	9,504	9,504	74,880/9,504
0.56	0.001	12,672	12,672	99,840/12,672
0.6446	0.001	15,840	15,840	124,800/15,840
0.8	0.001	120,640	120,640	957,696/120,640

Table 6.2. Parameter settings for infinite electrode simulations. Here "DoFs" stands for degrees of freedom.

ration, it cannot be sustained as in the slim electrode configuration. To substantiate this fact, we set $\mathcal{P} = 10$ and $\mathcal{R} = 800$ and compare the energies in Figure 6.13 for the two configurations. We observe that the energies in the slim case are not only significantly larger, they remain large as time evolves unlike in the infinite electrode configuration.



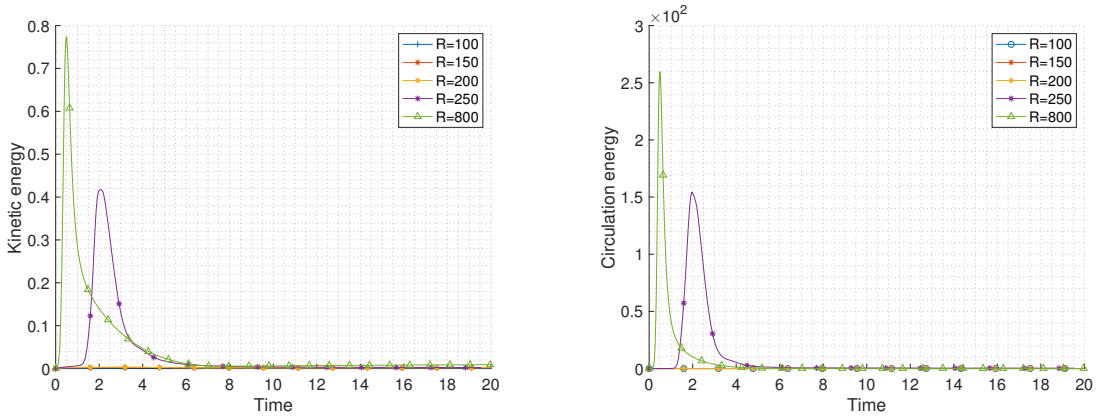
(a) (b)

Figure 6.11. Electroconvection for $\mathcal{P} = 10$, $\mathcal{R} = 100$, and $\alpha = 0.56$ at time $t = 40$. (a) Numerical approximation of the velocity field $\boldsymbol{u}$; (b) Numerical approximation of the electric surface charge density distribution q.

α	τ	M	DoFs for $\Psi_{0,\Omega}$	DoFs for q	DoFs for $\boldsymbol{u}$ and p
≤ 0.452	0.001	3	9,009	6,272	49,664/6,272
0.56, 0.6446	0.001	3	24,960	24,960	198,144/24,960
0.8	0.001	3	99,072	99,072	789,504/99,072

Table 6.3. Parameter settings for slim electrode simulations. Here "DoFs" stands for degrees of freedom.

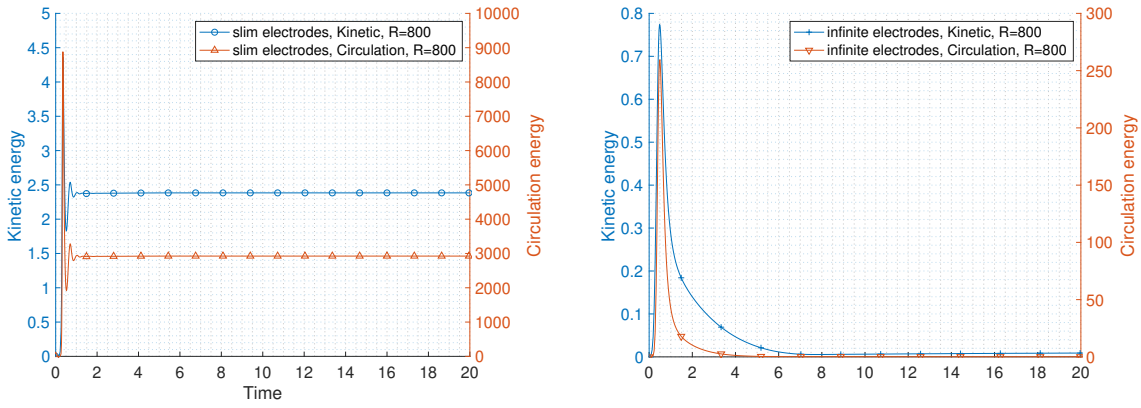
The difference between the two models is strikingly significant. In the presence of infinite electrodes model, electroconvection occurs as well but requires much larger Rayleigh number $\mathcal{R}$. Still, even when convective flows appears, they do not persist and quickly disappear as indicated by energies increasing significantly in the early stage of the process but eventually vanishing as the time evolves, see Figure 6.13. In contrast, the capability of maintaining a stable electroconvection phenomena indicates that the slim electrodes model is more adequate for electroconvection. Therefore, from now on we only consider the slim electrodes case. Worth mentioning, both mod-



(a) (b)

Figure 6.12. Comparison of the Kinetic energy (a) and circulation enegry (b) for the infinite electrode model with $\mathcal{P} = 10$, $\alpha = 0.33$ and for several values of $\mathcal{R}$.

els predict the same the number of pairs of vortices $\mathcal{N}_c$, which seems indicating that $\mathcal{N}_c$ depends mainly on the geometry as discussed later in Subsection 6.5.3.

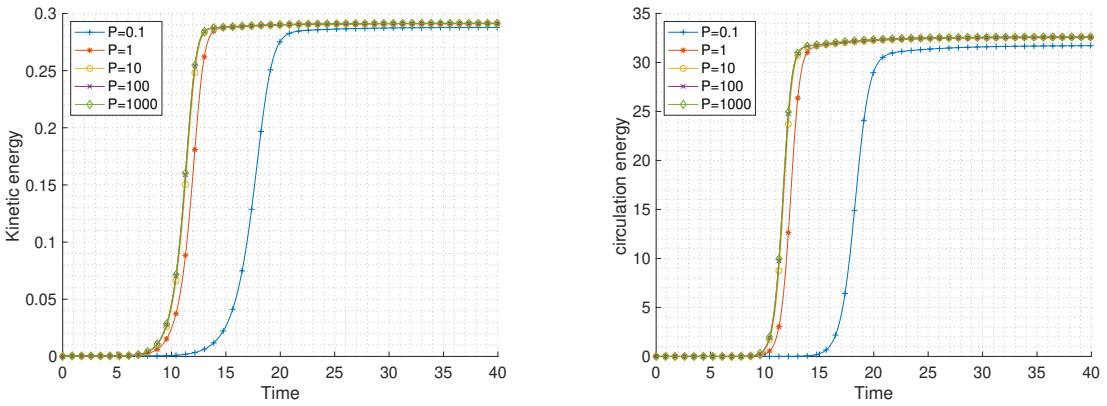


(a) (b)

Figure 6.13. Comparison of the evolution the energies E_k and E_{curl} versus time for the slim and infinite electrodes configurations with $\alpha = 0.33$, $\mathcal{R} = 800$ and $\mathcal{P} = 10$. The slim configuration energies are significantly larger and remain large during the entire evolution indicating a sustained electroconvection phenomena.

6.5.2 Effect of the Prandtl number for the slim electrodes case

The Prandtl number $\mathcal{P}$ is the dimensionless ratio between the charge and viscous relaxation times. To understand its influence in the electroconvection phenomena, we fix $\alpha = 0.33$, $\mathcal{R} = 100$ and let $\mathcal{P}$ vary from 0.1 to $1,000$. In Figure 6.14 we report the evolotion of kinetic energies and the circulation energies of the fluid. It appears that the value of $\mathcal{P}$ does not influence whether electroconvection occurs or not but rather its activation time. Eventually the long term behavior are the same for all simulations. Worth noting, the energy plots in Figure 6.14 seem also to indicate that electroconvection cannot occur before $t = 5$ irrespectively of the Prandtl number.



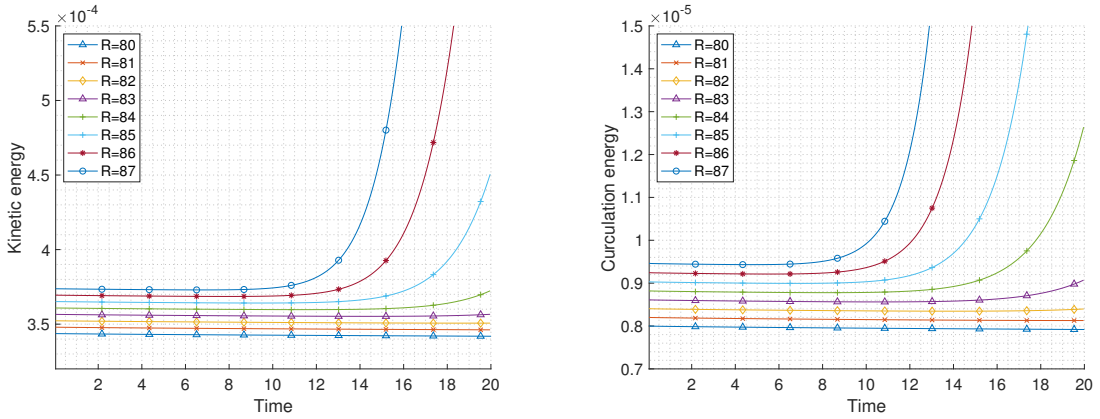
(a) (b)

Figure 6.14. Kinetic energy (a) and E_{curl} (b) evolutions over time for various Prandtl number with $\alpha = 0.33$ and $\mathcal{R} = 100$. Increasing the Prandtl number decreases the time for the electroconvection to develop. We observe that electroconvection cannot occur in this setting before $t = 5$ irrespectively of the Prandtl number. Compare with Figure 6.17.

6.5.3 Effect of the geometry for the slim electrodes case

The geometry of the annulus domain Ω is characterized by the aspect ratio $\alpha = R_i/R_o$; see (6.18). It turns out α affects the critical Rayleigh number $\mathcal{R}_c$ after which electroconvection occurs as well as the number of vortex pairs $\mathcal{N}_c$.

The observation in Subsection 6.5.2 ensures that we can fix $\mathcal{P} = 10$ and determine for various aspect ratios $\alpha \in [0.1, 0.8]$ the corresponding critical Rayleigh number $\mathcal{R}_c$. The latter is determined by running independent simulations starting from a relatively small $\mathcal{R}$ at which electroconvection does not occur, and continuously increasing the value $\mathcal{R}$ by one unit each time until electroconvection occurs. This allows us to determine the critical value $\mathcal{R}_c$ up to 1 unit. The system is considered steady when E_k and E_{curl} remains within 0.1% relative difference throughout the simulation time. For example, Figure 6.15 depicts the evolution of E_k and E_{curl} for $\alpha = 0.33$ with $\mathcal{R}$ varying from 80 to 87. The critical Rayleigh number satisfies $82 < \mathcal{R}_c < 83$.



(a) (b)

Figure 6.15. Kinetic energy (a) and circulation energy (b) for slim electrodes with $\mathcal{P} = 10$, $\alpha = 0.33$ and different values of $\mathcal{R}$.

In Figure 6.16(a) we report all the critical Rayleigh values for various α and compare them with [6]. Without completely matching, we consider our results in good agreement.

We shall see from a comparison between Figures 6.10 and 6.11 that the numbers of vortex pairs $\mathcal{N}_c$ are strongly influenced by the geometry. To determine the precise number of pairs we fix $\mathcal{P} = 10$ and set, for each aspect ratio considered, the Rayleigh number at the critical value $\mathcal{R}_c$. In Figure 6.16(b) we compare the number of vortex pairs $\mathcal{N}_c$ obtained by our algorithm with

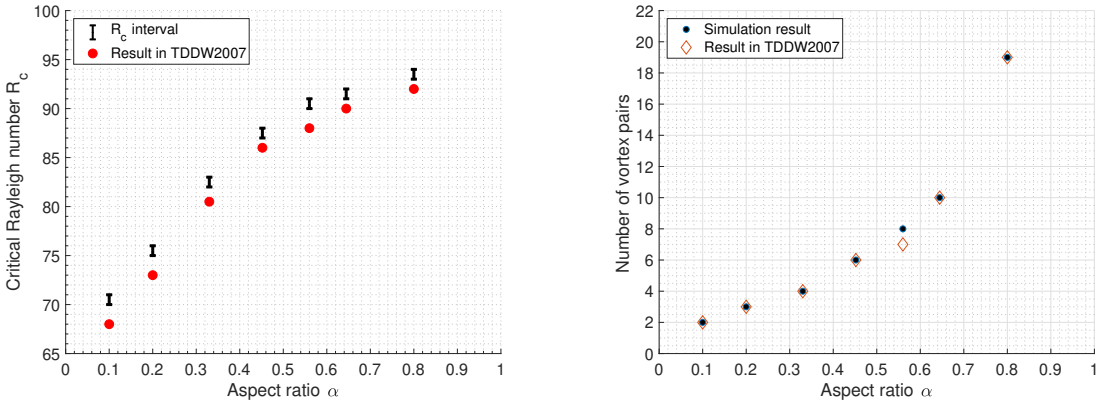


Figure 6.16. Critical Rayleigh number (a) and number of vortex pairs (b) in the slim electrode case for $\mathcal{P} = 10$ and several aspect ratios α. Compared with the results provided in [6]. The uncertainty intervals with length one in part (a) are due to the increment used in the critical Rayleigh number exploration.

the results obtained from [6]. They match for all aspect ratios considered except for $\alpha = 0.56$ where they differ by one. However, our predicted number of vortex pair in this case matches the experimental data reported Figure 5(b) of [6].

6.5.4 Effect of the Rayleigh number for the slim electrodes case

From the definition (6.17), we realize that $\mathcal{R} \propto V^2$. Increasing the Rayleigh number corresponds to a stronger electric field and thus a stronger Lorentz force. In Subsection 6.5.3, we have already discussed the influence of the geometry on the critical Rayleigh value. We now set $\alpha = 0.33$ and $\mathcal{P} = 10$ and complete the investigation by increasing the value of $\mathcal{R}$ up to $\approx 10\mathcal{R}_c$. The values of E_k and E_{curl} are reported in Figure 6.17. We observe that larger Rayleigh numbers result in faster activation of the convection. This is similar as for the Prandtl number discussed in Subsection 6.5.2 but in this case there does not seem to be a limiting time before which electroconvection cannot occur. Consequently fluids with larger Rayleigh numbers develop a stable electroconvection at earlier times. The strength of convection is also stronger for larger Rayleigh number.

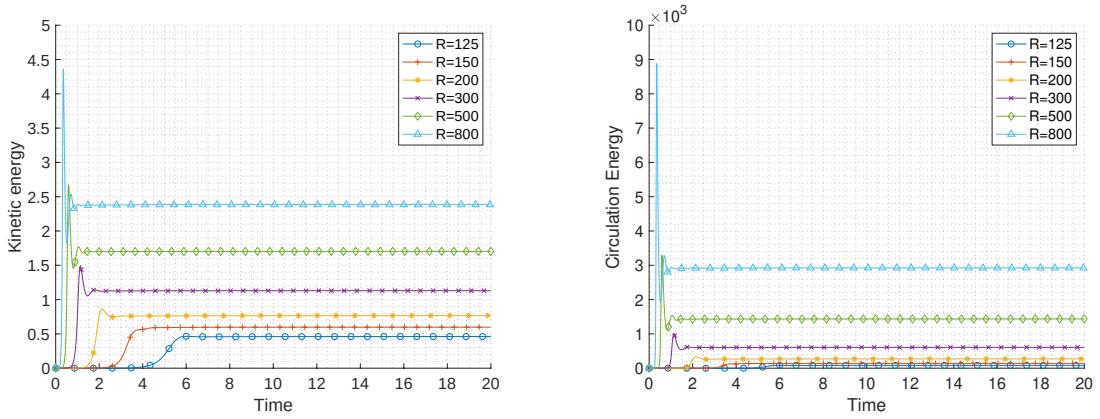


Figure 6.17. Comparison of the Kinetic energy (a) and circulation energy (b) for the slim electrode model with $\mathcal{P} = 10$, $\alpha = 0.33$ and for several values of $\mathcal{R}$.

7. SUMMARY AND CONCLUSIONS

This dissertation has provided numerical schemes for time-dependent problems with drift and involving fractional power of elliptic operators. The forward Euler time stepping was advocated for time discretization. On each time step, the stationary problem is approximated through sinc quadrature stepping to its Balakrishnan integral representation, and the finite element approximation. We remark that the resulting subproblems are mutually independent and only require solving standard diffusion-reaction problems, thus standard finite element software libraries can be readily applied. For the homogeneous problem, the L^2 error between the solution and its final approximation consists of three parts: the exponentially convergent sinc approximation, the first order time discretization, and the space discretization which only depends on the elliptic regularity index (see Assumption 2.2). The error analysis to the non-homogeneous problem was derived using results from homogeneous problems with the Duhamel's principle. It was worth mentioning that its space approximation error was also dependent on the smoothness of the right hand side data.

In Chapter 5 we have derived the governing equations for the surface quasi-geostrophic fluid dynamics from the conservation of potential temperature. The velocity equation indicates that the motion can only along the directions of iso-bars, and it requires solving another fractional Laplace problem. The numerical scheme utilized operator splitting, resulting in a parabolic equation involving fractional Laplacian and a standard transport equation. As a consequence, the numerical approximations to above equations only require the approximation techniques developed in Chapters 3 and 4. Two additional techniques were incorporated: the Strongly Stability Preserving two-stage Runge-Kutta (SSP-RK2) scheme to ensure second-order convergence in time, and the second-order smoothness-based maximum principle preserving artificial viscosity to eliminate spurious oscillations from continuous finite elements. For all simulations proposed in Section 5.5 our results were in good agreement with existing results.

In Chapter 6, we have extracted a mathematical model for electrically driven convection in an annular two dimensional fluid. Two different electrodes configurations are considered: infinite and

slim. Depending on the configurations, nonlocal representations of the electric potential are derived on the liquid domain. This together with the surface density charge conservation relation and the Navier-Stokes system for the fluid dynamics yield a system of partial differential equations defined only in the two dimensional and bounded liquid region. Our numerical algorithm took advantage of the proposed approximation schemes in Chapters 3 and 4. Our simulations reveal that the slim electrodes configuration is more favorable for electroconvection: it requires less energy and is able to sustain the effect. We therefore choose the slim configuration to provide a numerical study of the three nondimensional parameters describing the electroconvection system. We find that the Prandtl number does not affect the long-term behavior of the charge density distribution and the fluid dynamics. In opposition, for several electrodes aspect ratio, we provide critical Rayleigh numbers above which electroconvection occurs.

REFERENCES

[1] P. Constantin, M.-C. Lai, R. Sharma, Y.-H. Tseng, and J. Wu, "New numerical results for the surface quasi-geostrophic equation," *Journal of Scientific Computing*, vol. 50, no. 1, pp. 1–28, 2012.

[2] F. Song and G. E. Karniadakis, "Fractional spectral vanishing viscosity method: Application to the quasi-geostrophic equation," *Chaos, Solitons & Fractals*, vol. 102, pp. 327–332, 2017.

[3] V. Popa-Nita, I. Gerlič, and S. Kralj, "The influence of disorder on thermotropic nematic liquid crystals phase behavior," *International journal of molecular sciences*, vol. 10, no. 9, pp. 3971–4008, 2009.

[4] D. Porter, J. R. Savage, I. Cohen, P. Spicer, and M. Caggioni, "Temperature dependence of droplet breakup in 8cb and 5cb liquid crystals," *Physical Review E*, vol. 85, no. 4, p. 041701, 2012.

[5] S. Morris and P. Kruse, "Electroconvection in an unsheared smectic film." url = https://www.youtube.com/watch?v=aaKwymX7pqY, July 2010.

[6] P. Tsai, Z. A. Daya, V. B. Deyirmenjian, and S. W. Morris, "Direct numerical simulation of supercritical annular electroconvection," *Physical Review E*, vol. 76, no. 2, p. 026305, 2007.

[7] P. Carr, H. Geman, D. B. Madan, and M. Yor, "The fine structure of asset returns: An empirical investigation," *The Journal of Business*, vol. 75, no. 2, pp. 305–332, 2002.

[8] S. Levendorski, "Pricing of the american put under lévy processes," *International Journal of Theoretical and Applied Finance*, vol. 7, no. 03, pp. 303–335, 2004.

[9] W. Chen, "A speculative study of 2/3-order fractional laplacian modeling of turbulence: Some thoughts and conjectures," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 16, no. 2, p. 023126, 2006.

[10] L. Caffarelli and A. Vasseur, "Drift diffusion equations with fractional diffusion and the quasi-geostrophic equation," *arXiv preprint math/0608447*, 2006.

[11] P. Gatto and J. S. Hesthaven, "Numerical approximation of the fractional laplacian via finite

elements, with an application to image denoising," *Journal of Scientific Computing*, vol. 65, no. 1, pp. 249–270, 2015.

[12] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling & Simulation*, vol. 7, no. 3, pp. 1005–1028, 2008.

[13] I. M. Held, R. T. Pierrehumbert, S. T. Garner, and K. L. Swanson, "Surface quasi-geostrophic dynamics," *Journal of Fluid Mechanics*, vol. 282, pp. 1–20, 1995.

[14] R. Musina and A. I. Nazarov, "On fractional laplacians," *Communications in Partial Differential Equations*, vol. 39, no. 9, pp. 1780–1790, 2014.

[15] M. Ilic, F. Liu, I. Turner, and V. Anh, "Numerical approximation of a fractional-in-space diffusion equation, i," *Fractional Calculus and Applied Analysis*, vol. 8, no. 3, pp. 323–341, 2005.

[16] M. Ilic, F. Liu, I. Turner, and V. Anh, "Numerical approximation of a fractional-in-space diffusion equation (ii)–with nonhomogeneous boundary conditions," *Fractional Calculus and applied analysis*, vol. 9, no. 4, pp. 333–349, 2006.

[17] R. H. Nochetto, E. Otárola, and A. J. Salgado, "A pde approach to fractional diffusion in general domains: a priori error analysis," *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 733–791, 2015.

[18] R. H. Nochetto, E. Otárola, and A. J. Salgado, "A pde approach to fractional diffusion in general domains: a priori error analysis," *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 733–791, 2015.

[19] R. H. Nochetto, E. Otarola, and A. J. Salgado, "A pde approach to numerical fractional diffusion," *arXiv preprint arXiv:1508.04382*, 2015.

[20] L. Caffarelli and L. Silvestre, "An extension problem related to the fractional laplacian," *Communications in partial differential equations*, vol. 32, no. 8, pp. 1245–1260, 2007.

[21] P. R. Stinga and J. L. Torrea, "Extension problem and harnack's inequality for some fractional operators," *Communications in Partial Differential Equations*, vol. 35, no. 11, pp. 2092–2122, 2010.

[22] T. Kato, "Fractional powers of dissipative operators," *Journal of the Mathematical Society of Japan*, vol. 13, no. 3, pp. 246–274, 1961.

[23] A. Bonito and J. Pasciak, "Numerical approximation of fractional powers of elliptic operators," *Mathematics of Computation*, vol. 84, no. 295, pp. 2083–2110, 2015.

[24] J. Lund and K. L. Bowers, *Sinc methods for quadrature and differential equations*. SIAM, 1992.

[25] A. Bonito and J. E. Pasciak, "Numerical approximation of fractional powers of regularly accretive operators," *IMA Journal of Numerical Analysis*, p. drw042, 2016.

[26] A. Bonito, W. Lei, and J. E. Pasciak, "On sinc quadrature approximations of fractional powers of regularly accretive operators," *Journal of Numerical Mathematics*.

[27] A. Bonito, W. Lei, and J. E. Pasciak, "Numerical approximation of the integral fractional laplacian," *Numerische Mathematik*, pp. 1–44, 2017.

[28] A. Bonito, J. P. Borthagaray, R. H. Nochetto, E. Otarola, and A. J. Salgado, "Numerical methods for fractional diffusion," *Computing and Visualization in Science*, vol. 19, no. 5-6, pp. 19–46, 2018.

[29] A. Bonito, W. Lei, and J. E. Pasciak, "The approximation of parabolic equations involving fractional powers of elliptic operators," *Journal of Computational and Applied Mathematics*, vol. 315, pp. 32–48, 2017.

[30] V. Thomee, *Galerkin Finite Element Methods for Parabolic Problems*, vol. 25. Springer Science & Business Media, 2007.

[31] A. E. Gill, *Atmosphere-ocean dynamics*. Elsevier, 2016.

[32] A. Majda, *Introduction to PDEs and Waves for the Atmosphere and Ocean*, vol. 9. American Mathematical Soc., 2003.

[33] J. Pedlosky, *Geophysical fluid dynamics*. Springer Science & Business Media, 2013.

[34] P. Constantin, A. J. Majda, and E. Tabak, "Formation of strong fronts in the 2-d quasi-geostrophic thermal active scalar," *Nonlinearity*, vol. 7, no. 6, p. 1495, 1994.

[35] S. Gottlieb and L.-A. J. Gottlieb, "Strong stability preserving properties of runge–kutta time

discretization methods for linear constant coefficient operators," *Journal of Scientific Computing*, vol. 18, no. 1, pp. 83–109, 2003.

[36] J.-L. Guermond and B. Popov, "Invariant domains and second-order continuous finite element approximation for scalar conservation equations," *SIAM Journal on Numerical Analysis*, vol. 55, no. 6, pp. 3120–3146, 2017.

[37] P.-C. Tsai, *The Route to Chaos and Turbulence in Annular Electroconvection*. University of Toronto, 2007.

[38] R. A. Adams and J. J. Fournier, *Sobolev spaces*, vol. 140. Elsevier, 2003.

[39] L. Tartar, *An introduction to Sobolev spaces and interpolation spaces*, vol. 3. Springer Science & Business Media, 2007.

[40] J. H. Bramble and X. Zhang, "The analysis of multigrid methods," *HandBOOK of numerical analysis*, vol. 7, pp. 173–415, 2000.

[41] T. Kato, "Fractional powers of dissipative operators," *Journal of the Mathematical Society of Japan*, vol. 13, no. 3, pp. 246–274, 1961.

[42] W. Lei, "Numerical approximation of partial differential equations involving fractional differential operators," 2018.

[43] G. B. Folland, *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

[44] A. Lunardi, *Interpolation theory*. Edizioni della normale, 2009.

[45] J. G. Charney, "On the scale of atmospheric motions," 1948.

[46] J. LaCasce and A. Mahadevan, "Estimating subsurface horizontal and vertical velocities from sea-surface temperature," *Journal of Marine Research*, vol. 64, no. 5, pp. 695–721, 2006.

[47] G. Lapeyre and P. Klein, "Dynamics of the upper oceanic layers in terms of surface quasigeostrophy theory," *Journal of physical oceanography*, vol. 36, no. 2, pp. 165–176, 2006.

[48] G. K. Vallis, *Atmospheric and oceanic fluid dynamics*. Cambridge University Press, 2017.

[49] B. J. Hoskins, "The geostrophic momentum approximation and the semi-geostrophic equations," *Journal of the Atmospheric Sciences*, vol. 32, no. 2, pp. 233–242, 1975.

[50] B. J. Hoskins and F. P. Bretherton, "Atmospheric frontogenesis models: Mathematical formulation and solution," *Journal of the Atmospheric Sciences*, vol. 29, no. 1, pp. 11–37, 1972.

[51] C. Leith, "Nonlinear normal mode initialization and quasi-geostrophic theory," *Journal of the Atmospheric Sciences*, vol. 37, no. 5, pp. 958–968, 1980.

[52] X. Capet, P. Klein, B. L. Hua, G. Lapeyre, and J. C. Mcwilliams, "Surface kinetic energy transfer in surface quasi-geostrophic flows," *Journal of Fluid Mechanics*, vol. 604, pp. 165–174, 2008.

[53] J. G. Charney, "Geostrophic turbulence," *Journal of the Atmospheric Sciences*, vol. 28, no. 6, pp. 1087–1095, 1971.

[54] G. Lapeyre, "Surface quasi-geostrophy," *Fluids*, vol. 2, no. 1, p. 7, 2017.

[55] P. Constantin, Q. Nie, and N. Schörghofer, "Nonsingular surface quasi-geostrophic flow," *Physics Letters A*, vol. 241, no. 3, pp. 168–172, 1998.

[56] R. Glowinski, "Finite element methods for incompressible viscous flow," *Handbook of numerical analysis*, vol. 9, pp. 3–1176, 2003.

[57] J.-L. Guermond, M. Q. de Luna, B. Popov, C. E. Kees, and M. W. Farthing, "Well-balanced second-order finite element approximation of the shallow water equations with friction," *SIAM Journal on Scientific Computing*, vol. 40, no. 6, pp. A3873–A3901, 2018.

[58] K. Ohkitani and M. Yamada, "Inviscid and inviscid-limit behavior of a surface quasi-geostrophic flow," *Physics of Fluids*, vol. 9, no. 4, pp. 876–882, 1997.

[59] E. Bodenschatz, W. Pesch, and G. Ahlers, "Recent developments in rayleigh-bénard convection," *Annual review of fluid mechanics*, vol. 32, no. 1, pp. 709–778, 2000.

[60] S. Rosenblat, S. Davis, and G. Homsy, "Nonlinear marangoni convection in bounded layers. part 1. circular cylindrical containers," *Journal of fluid Mechanics*, vol. 120, pp. 91–122, 1982.

[61] S. Rosenblat, G. Homsy, and S. Davis, "Nonlinear marangoni convection in bounded layers. part 2. rectangular cylindrical containers," *Journal of Fluid Mechanics*, vol. 120, pp. 123–138, 1982.

[62] R. G. Larson, "Instabilities in viscoelastic flows," *Rheologica Acta*, vol. 31, no. 3, pp. 213–263, 1992.

[63] D. Saville, "Electrohydrodynamics: the taylor-melcher leaky dielectric model," *Annual review of fluid mechanics*, vol. 29, no. 1, pp. 27–64, 1997.

[64] A. Ciferri, *Polymer liquid crystals*. Elsevier, 2012.

[65] A. Buka and L. Kramer, *Pattern formation in liquid crystals*. Springer Science & Business Media, 2012.

[66] S. W. Morris, J. R. De Bruyn, and A. May, "Patterns at the onset of electroconvection in freely suspended smectic films," *Journal of statistical physics*, vol. 64, no. 5-6, pp. 1025–1043, 1991.

[67] Z. A. Daya, S. W. Morris, and J. R. De Bruyn, "Electroconvection in a suspended fluid film: a linear stability analysis," *Physical Review E*, vol. 55, no. 3, p. 2682, 1997.

[68] V. Deyirmenjian, Z. A. Daya, and S. W. Morris, "Weakly nonlinear analysis of electroconvection in a suspended fluid film," *Physical Review E*, vol. 56, no. 2, p. 1706, 1997.

[69] V. Deyirmenjian, Z. A. Daya, and S. W. Morris, "Codimension-two points in annular electroconvection as a function of aspect ratio," *Physical Review E*, vol. 72, no. 3, p. 036211, 2005.

[70] Z. A. Daya, V. Deyirmenjian, S. W. Morris, and J. R. De Bruyn, "Annular electroconvection with shear," *Physical review letters*, vol. 80, no. 5, p. 964, 1998.

[71] Z. A. Daya, V. Deyirmenjian, and S. W. Morris, "Electrically driven convection in a thin annular film undergoing circular couette flow," *Physics of Fluids*, vol. 11, no. 12, pp. 3613–3628, 1999.

[72] Z. A. Daya, V. Deyirmenjian, and S. W. Morris, "Sequential bifurcations in sheared annular electroconvection," *Physical Review E*, vol. 66, no. 1, p. 015201, 2002.

[73] P. Tsai, Z. A. Daya, and S. W. Morris, "Aspect-ratio dependence of charge transport in turbulent electroconvection," *Physical review letters*, vol. 92, no. 8, p. 084503, 2004.

[74] S. J. Kim, Y.-A. Song, and J. Han, "Nanofluidic concentration devices for biomolecules uti-

lizing ion concentration polarization: theory, fabrication, and applications," *Chemical Society Reviews*, vol. 39, no. 3, pp. 912–922, 2010.

[75] R. Kwak, V. S. Pham, K. M. Lim, and J. Han, "Shear flow of an electrically charged fluid by ion concentration polarization: Scaling laws for electroconvective vortices," *Physical Review Letters*, vol. 110, no. 11, 2013.

[76] P. Constantin, T. Elgindi, M. Ignatova, and V. Vicol, "On some electroconvection models," *Journal of Nonlinear Science*, pp. 1–15, 2016.

[77] D. J. Griffiths, "Introduction to electrodynamics," 2005.

[78] H. Antil, J. Pfefferer, and S. Rogovs, "Fractional operators with inhomogeneous boundary conditions: Analysis, control, and discretization," *arXiv preprint arXiv:1703.05256*, 2017.

[79] L. C. Evans, *Partial differential equations*. American Mathematical Society, 2010.

[80] G. B. Arfken and H. J. Weber, "Mathematical methods for physicists," 1999.

[81] A. Prudnikov, A. Brychkov Yu, and O. Marichev, *Integrals and Series vol 2 (New York: Gordon and Breach)*. 1990.

[82] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Academic press, 2014.

[83] M. Abramowitz and I. A. Stegun, *HandBOOK of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55. Courier Corporation, 1964.

[84] G. Alzetta, D. Arndt, W. Bangerth, V. Boddu, B. Brands, D. Davydov, R. Gassmoeller, T. Heister, L. Heltai, K. Kormann, M. Kronbichler, M. Maier, J.-P. Pelteret, B. Turcksin, and D. Wells, "The `deal.II` library, version 9.0," *Journal of Numerical Mathematics*, vol. 26, no. 4, pp. 173–183, 2018.

[85] S. Brenner and R. Scott, *The mathematical theory of finite element methods*, vol. 15. Springer Science & Business Media, 2007.

[86] C. Brändle, E. Colorado, A. de Pablo, and U. Sánchez, "A concave-convex elliptic problem involving the fractional laplacian," *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, vol. 143, no. 1, pp. 39–71, 2013.

APPENDIX A

A VARIATION OF CAFFARELLI-SILVESTRE EXTENSION

Consider the following extended problem: find $\mathscr{U}(\boldsymbol{x}, y) : \mathbb{R}^d \times [0, \infty) \to \mathbb{R}$ satisfying

$$\nabla\cdot(y^\alpha \nabla \mathscr{U}) = 0 \quad \text{in } \mathbb{R}^d \times (0, \infty), \tag{A.1}$$

with boundary conditions

$$y^\alpha \partial_y \mathscr{U} = q \quad \text{on } \Omega, \quad \text{and} \quad \mathscr{U} = 0 \quad \text{on } \mathbb{R}^d \backslash \Omega, \tag{A.2}$$

and closed with the vanishing condition

$$\lim_{y\to\infty} \mathscr{U} = 0. \tag{A.3}$$

Here the constant $\alpha \in (-1, 1)$, and the function $q \in L^2(\Omega)$.

Also consider the following fractional problem: for a fractional power $s = \frac{1-\alpha}{2} \in (0, 1)$, seek $u \in D((-\Delta_{\mathscr{F}})^s)$, such that

$$(-\Delta_{\mathscr{F}})^s u(x) = \frac{q}{D_s} \quad \text{on } \Omega, \quad \text{and} \quad u = 0 \quad \text{on } \mathbb{R}^d \backslash \Omega. \tag{A.4}$$

Here the fractional operator is of integral type defined by (1.2), and the constant

$$D_s = 2^{1-2s}\Gamma(1-s)/\Gamma(s) = 2^\alpha \Gamma(\frac{1+\alpha}{2})/\Gamma(\frac{1-\alpha}{2}).$$

In this appendix we will show that $\mathscr{U}$ restricted to $\mathbb{R}^d$ equals u by examining their corresponding energy functionals coincide. We therefore first derive the energy functionals independently.

Lemma A.1. The energy functional to (A.1), (A.2), and (A.3) is

$$I[\mathscr{W}] = \frac{1}{2}\int_{\mathbb{R}^d}\int_0^\infty |\nabla \mathscr{W}|^2\, y^\alpha \, dy\, d\boldsymbol{x} - \int_\Omega q\mathscr{W}(\boldsymbol{x},0)\, d\boldsymbol{x}.$$

Here $\mathscr{W}$ belongs to the admissible set

$$\mathcal{A} := \{\mathscr{W} \in H^1(\mathbb{R}^d \times \mathbb{R}_+) \ : \ \mathscr{W} = 0 \text{ on } \mathbb{R}^d\backslash\Omega, \ \lim_{y\to\infty} \mathscr{W} = 0\}.$$

In addition, the solution $\mathscr{U}$ to (A.1), (A.2), and (A.3) minimizes $I[\mathscr{W}]$ for $\mathscr{W} \in \mathcal{A}$. Conversely, the minimizer to $I[\mathscr{W}]$ is a solution to (A.1), (A.2), and (A.3).

Proof. 1 For an arbitrary $\mathscr{W} \in \mathcal{A}$ an integration by part yields

$$\begin{aligned} 0 &= \int_{\mathbb{R}^d}\int_0^\infty -\nabla\cdot(y^\alpha \nabla\mathscr{U})(\mathscr{U} - \mathscr{W})\, dy\, d\boldsymbol{x} \\ &= \int_{\mathbb{R}^d}\int_0^\infty y^\alpha \nabla\mathscr{U}\nabla(\mathscr{U} - \mathscr{W})\, dy\, d\boldsymbol{x} - \int_\Omega q(\mathscr{U} - \mathscr{W})\, d\boldsymbol{x}. \end{aligned}$$

Employing a Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} \int_{\mathbb{R}^d}\int_0^\infty y^\alpha |\nabla\mathscr{U}|^2 \, dy\, d\boldsymbol{x} - \int_\Omega q\mathscr{U}\, d\boldsymbol{x} \le \frac{1}{2}\int_{\mathbb{R}^d}\int_0^\infty y^\alpha |\nabla\mathscr{U}|^2 \, dy\, d\boldsymbol{x} \\ + \frac{1}{2}\int_{\mathbb{R}^d}\int_0^\infty y^\alpha |\nabla\mathscr{W}|^2 \, dy\, d\boldsymbol{x} - \int_\Omega q\mathscr{W}\, d\boldsymbol{x}. \end{aligned}$$

This shows that $\mathscr{U}$ is indeed a minimizer to $I[\mathscr{W}]$.

2 Now suppose $\mathscr{U}$ is the minimizer to $I[\mathscr{W}]$, we then introduce a function

$$f(t) := I[\mathscr{U} + t\mathscr{W}], \quad \text{for } t \in \mathbb{R}.$$

Expanding $f(t)$ gives

$$f(t) = \frac{1}{2}\int_{\mathbb{R}^d}\int_0^\infty y^\alpha |\nabla\mathscr{U}|^2 + 2ty^\alpha\nabla\mathscr{U}\cdot\nabla\mathscr{W} + t^2 y^\alpha |\nabla\mathscr{W}|^2 \, dy\, d\boldsymbol{x} - \int_\Omega q(\mathscr{U} + t\mathscr{W})\, d\boldsymbol{x}.$$

Then in view of the minimizing property of $\mathscr{U}$, we have $f'(0) = 0$, which means for any $\mathscr{W} \in \mathcal{A}$,

$$0 = \int_{\mathbb{R}^d} \int_0^\infty y^\alpha \nabla \mathscr{U} \cdot \nabla \mathscr{W} \, dy \, d\boldsymbol{x} - \int_\Omega q \mathscr{W} \, d\boldsymbol{x} = \int_{\mathbb{R}^d} \int_0^\infty -\nabla \cdot (y^\alpha \nabla \mathscr{U}) \mathscr{W} \, dy \, d\boldsymbol{x}.$$

Here the last equality invoked the integrate by part and boundary conditions (A.2), and (A.3). This identity clearly implies (A.1). □

Lemma A.2. The energy functional to (A.4) is

$$J[w] = \frac{1}{2} \int_{\mathbb{R}^d} |\xi|^{2s} \left|\hat{w}(\xi)\right|^2 \, d\xi - \frac{1}{D_s} \int_\Omega q w \, d\boldsymbol{x},$$

where $\hat{\cdot}$ denotes the Fourier transform defined on $\mathbb{R}^d$. Here w belongs to the admissible set

$$\mathcal{B} = \{w \in H^1(\mathbb{R}^d) : w = 0 \text{ on } \mathbb{R}^d \backslash \Omega\}.$$

In addition, the solution u to (A.4) minimizes $J[w]$ for $w \in \mathcal{B}$. Conversely, the minimizer to $J[w]$ is a solution to (A.4).

Proof. 1 On one hand, choose $w \in \mathcal{B}$, which allows us to do an integrate by part to obtain

$$\begin{aligned} 0 &= \int_{\mathbb{R}^d} ((-\Delta_{\mathscr{F}})^s u - \frac{q}{D_s})(u - w) \, d\boldsymbol{x} \\ &= \int_{\mathbb{R}^d} (-\Delta_{\mathscr{F}})^s u u \, d\boldsymbol{x} - \int_{\mathbb{R}^d} (-\Delta_{\mathscr{F}})^s u w \, d\boldsymbol{x} - \int_{\mathbb{R}^d} \frac{q}{D_s} u \, d\boldsymbol{x} + \int_{\mathbb{R}^d} \frac{q}{D_s} w \, d\boldsymbol{x} \end{aligned}$$

Notice that by [27, Theorem 4.1], the fractional powers can be balanced between the two terms:

$$\begin{aligned} \int_{\mathbb{R}^d} (-\Delta_{\mathscr{F}})^s u u \, d\boldsymbol{x} &= \int_{\mathbb{R}^d} \left|(-\Delta_{\mathscr{F}})^{s/2} u\right|^2 \, d\boldsymbol{x}, \\ \int_{\mathbb{R}^d} (-\Delta_{\mathscr{F}})^s u w \, d\boldsymbol{x} &= \int_{\mathbb{R}^d} (-\Delta_{\mathscr{F}})^{s/2} u (-\Delta_{\mathscr{F}})^{s/2} w \, d\boldsymbol{x} \\ &\leq \frac{1}{2} \int_{\mathbb{R}^d} \left|(-\Delta_{\mathscr{F}})^{s/2} u\right|^2 + \left|(-\Delta_{\mathscr{F}})^{s/2} w\right|^2 \, d\boldsymbol{x} \end{aligned}$$

Here the last inequality utilized Cauchy-Schwatz inequality. After applying Parserval's identity

and the definition of integral fractional Laplacian, we can show that

$$\int_{\mathbb{R}^d} \left|(-\Delta_{\mathscr{F}})^{s/2} u\right|^2 \, d\boldsymbol{x} = \int_{\mathbb{R}^d} |\xi|^{2s} \, |\hat{u}(\xi)|^2 \, d\xi.$$

The same equality holds for u replaced by w. Putting them together, we obtain

$$\frac{1}{2}\int_{\mathbb{R}^d} |\xi|^{2s} \, |\hat{u}(\xi)|^2 \, d\xi - \int_{\mathbb{R}^d} \frac{q}{D_s} u \, d\boldsymbol{x} \leq \frac{1}{2}\int_{\mathbb{R}^d} |\xi|^{2s} \, |\hat{w}(\xi)|^2 \, d\xi - \int_{\mathbb{R}^d} \frac{q}{D_s} w \, d\boldsymbol{x}.$$

2 Now suppose $u = \arg\min J[w]$. For any $w \in \mathcal{B}$, the function

$$g(t) := J[u + tw] = \frac{1}{2}\int_{\mathbb{R}^d} |\xi|^{2s} \, |\hat{u} + t\hat{w}|^2 \, d\xi - \int_{\Omega} \frac{q}{D_s}(u + tw) \, d\boldsymbol{x}$$

admits minimum at $t = 0$, hence $g'(0) = 0$, namely,

$$0 = \int_{\mathbb{R}^d} |\xi|^{2s} \, \hat{u}(\xi)\hat{w}(\xi) \, d\xi - \int_{\Omega} \frac{q}{D_s} w \, d\boldsymbol{x} = \int_{\Omega} \left((-\Delta_{\mathscr{F}})^s u - \frac{q}{D_s} \right) w \, d\boldsymbol{x}.$$

Since $w \in \mathcal{B}$ is arbitrary, we have

$$(-\Delta_{\mathscr{F}})^s u = \frac{q}{D_s}.$$

This is exactly (A.4). □

We are now in position to derive the equivalence relation between the solution of the two problems.

Theorem A.3. Let $\mathscr{U}$ be the solution to (A.1) and u be the solution to (A.4). Then

$$\mathscr{U}(x, 0) = u(x).$$

Proof. The proof in essence follows [20, Section 3.2].

1 suppose $\mathscr{W}$ satisfies

$$\nabla \cdot (y^\alpha \nabla \mathscr{W}) = 0 \text{ in } \mathbb{R}^d \times (0, \infty), \quad \text{and } \lim_{y \to \infty} \mathscr{W} = 0.$$

Then equivalently $\mathscr{W}$ satisfies

$$\Delta_{\boldsymbol{x}} \mathscr{W} + \frac{\alpha}{y} \mathscr{W}_y + \mathscr{W}_{yy} = 0.$$

After applying Fourier transform on $\mathbb{R}^d$ we obtain

$$-|\xi|^2 \hat{\mathscr{W}}(\xi, y) + \frac{\alpha}{y} \hat{\mathscr{W}}_y(\xi, y) + \hat{\mathscr{W}}_{yy}(\xi, y) = 0, \quad \text{and } \lim_{y \to \infty} \hat{\mathscr{W}}(\xi, y) = 0.$$

This is an ordinary differential equation for each fixed ξ, and has a unique solution

$$\hat{\mathscr{W}}(\xi, y) = \hat{\mathscr{W}}(\xi, 0) \phi(|\xi|\, y).$$

Here the function $\phi(x)$ is the solution to the following equation:

$$-y^\alpha \phi(y) + (y^\alpha \phi'(y))' = 0, \qquad \phi(0) = 1, \qquad \lim_{y \to \infty} \phi(y) = 0.$$

It is not difficult to derive that it corresponds to energy functional

$$K[\psi] := \int_0^\infty (|\psi'|^2 + |\psi|^2) y^\alpha \, dy$$

associated with admissible set

$$\mathcal{C} = \{\psi : \psi(0) = 1\}.$$

Furthermore, it is shown in [86] that the above equation admits a solution ϕ to be a combination of

Bessel functions, and satisfies the following asymptotic properties:

$$\lim_{y\to 0}\phi(y)=1-\frac{2^{\alpha}\Gamma(\frac{1+\alpha}{2})}{(1-\alpha)\Gamma(\frac{1-\alpha}{2})}y^{1-\alpha},\quad \text{and } \lim_{y\to\infty}\phi(y)=\frac{\sqrt{2^{\alpha}\pi}}{\Gamma(\frac{1-\alpha}{2})}y^{-\alpha/2}e^{-y}.$$

We thus obtain the explicit value of $K[\phi]$, namely

$$K[\phi]=-(y^{\alpha}\phi'(y))\Big|_{y=0}^{y=\infty}=\frac{2^{\alpha}\Gamma(\frac{1+\alpha}{2})}{\Gamma(\frac{1-\alpha}{2})}=D_s.$$

Therefore,

$$\begin{aligned}\int_{\mathbb{R}^d}\int_0^{\infty}y^{\alpha}\left|\nabla\mathscr{W}\right|^2\,dy\,d\boldsymbol{x}&=\int_{\mathbb{R}^d}\int_0^{\infty}y^{\alpha}\left(\left|\nabla_{\boldsymbol{x}}\mathscr{W}\right|^2+\left|\mathscr{W}_y\right|^2\right)\,dy\,d\boldsymbol{x}\\&=\int_{\mathbb{R}^d}\int_0^{\infty}y^{\alpha}\left(|\xi|^2\left|\hat{\mathscr{W}}\right|^2+\left|\hat{\mathscr{W}}_y\right|^2\right)\,dy\,d\xi\\&=\int_{\mathbb{R}^d}|\xi|^{1-\alpha}\left|\hat{\mathscr{W}}(\xi,0)\right|^2\int_0^{\infty}(|\xi|\,y)^{\alpha}\left(|\phi(|\xi|\,y)|^2+|\phi'(|\xi|\,y)|^2\right)\,d(\xi y)\,d\xi\\&=D_s\int_{\mathbb{R}^d}|\xi|^{2s}\left|\hat{\mathscr{W}}(\xi,0)\right|^2\,d\xi.\end{aligned}$$

2 On one hand, $\mathscr{U}(x,0)\in\mathcal{B}$, thus

$$J[\mathscr{U}(x,0)]\geq J[u].$$

On the other hand, for any $w\in\mathcal{B}$, its extension $\tilde{w}$ such that $\nabla\cdot(y^{\alpha}\nabla\tilde{w})=0$ and $\lim_{y\to\infty}\tilde{w}=0$ is in $\mathcal{A}$, and satisfies

$$D_sJ[w]=I[\tilde{w}]\geq I[\mathscr{U}]=D_sJ[\mathscr{U}(x,0)],$$

Therefore $\mathscr{U}(x,0)$ is a minimizer to the functional $J[w]$, but since $u(x)$ is already an minimizer to $J[w]$, we derive that

$$\mathscr{U}(x,0)=u(x).$$

The proof is complete. □